BUS. ADM.
LIBRARY

JUN 27 1949

Journal of the

AMERICAN STATISTICAL ASSOCIATION

JUNE 1949

The Current Status of State and Local Population Estimates in the Census Bureau
Henry S. Shryock, Jr. and Norman Lawrence 157

The Uses and Usefulness of Binomial Probability Paper
Frederick Mosteller and John W. Tukey 174

Teaching Statistical Quality Control for Town and Gown
Edwin G. Olds and Lloyd A. Knowler 213

The Use of Sampling in Great Britain *C. A. Moser* 231

Unemployment and Migration in the Depression (1930-1935)
Ronald Freedman and Amos H. Hawley 260

Minimum X^2 and Maximum Likelihood Solution in Terms of a Linear Transform, with Particular Reference to Bio-Assay *Joseph Berkson* 273

Some Inadequacies of the Federal Censuses of Agriculture *Raymond J. Jessen* 279

The Edge Marking of Statistical Cards *A. M. Lester* 293

Conrad Alexander Verrijn Stuart (1865-1948) *Walter F. Willcox* 295

Proceedings of the 108th Annual Meeting 297

BOOK REVIEWS by T. N. E. Greville, Margaret J. Hagood, Burton H. Camp, Alexander M. Mood, Joseph Zubin, Robert C. Myers, D. B. DeLury and Boyd Harshbarger, Sebastian B. Littauer, Charles R. Scott, Jr., T. W. Anderson, Benjamin Epstein, F. G. Cornell, G. A. Baker, J. Wolfowitz, H. Burke Horton, Gregor Sebba

VOLUME 44 NUMBER 246

CHAPTER SECRETARIES

ALBANY	Murray Dorkin, *40 Glenwood Street, Albany, New York*
ASSOCIACIÓN CUBANA DE ESTADISTICA	Hugo Vivo, *Calle 28 No. 270, Apt. B, entre 21y 23, Vedado, Habana, Cuba*
AUSTIN	John R. Stockton, *113 Waggener Hall, University of Texas, Austin, Texas*
BOSTON	Wilfred S. Lake, *Northeastern University, Boston 15, Massachusetts*
CENTRAL INDIANA	Miss Lucy Kantz, *1504 N. Penn. 45, Indianapolis 2, Indiana*
CENTRAL NEW JERSEY	William Netschert, Jr., *1 Wayside Ave., Lawrence Township, Rt. 101, Trenton, N. J.*
CHICAGO	Randolph T. Snively, *Illinois Bell Telephone Company, Room 1902, 208 West Washington St., Chicago 6, Illinois*
CINCINNATI	Lee M. Welsh, *Procter & Gamble, Gwynne Bldg., Cincinnati 1, Ohio*
CLEVELAND	Gale R. Ober, Jr., *76 South Main St., Chagrin Falls, Ohio*
COLUMBUS	Mikhail V. Condoide, *188 West 10th Avenue, Columbus 1, Ohio*
CONNECTICUT	David F. Votaw, *Yale University, New Haven, Connecticut*
DENVER	Miss Lucy M. Klein, *1205 Washington St., Apt. 22, Denver 3, Colorado*
DETROIT	Andrew T. Court, *1517 Iroquois, Detroit 11, Michigan*
HARRISBURG STATISTICAL SOCIETY	Harry Hoyle, *1824 North 4th St., Harrisburg, Pennsylvania*
HAWAII	Margaret S. Wright, *P. O. Box 1222, Honolulu, T.H.*
ILLINOIS, UNIVERSITY OF	Donald W. Paden, *Department of Economics, University of Illinois, Urbana, Illinois*
ITHACA	Harry J. Loberg, *College of Engineering, Cornell University, Ithaca, New York*
LOS ANGELES	William B. Rice, *Box 3519, Terminal Annex, Los Angeles, California*
MADISON, WISCONSIN	Philip J. Fox, *403 Sterling Hall, University of Wisconsin, Madison, Wisconsin*
NEW YORK	Robert E. Lewis, *Federal Reserve Bank of New York, 33 Liberty St., New York 7, New York*
NORTH CAROLINA	George Nicholson, *Department of Mathematical Statistics, University of North Carolina*
OKLAHOMA CITY	Robert S. McCauley, *1710 Exchange, Oklahoma City, Oklahoma*
PACIFIC NORTHWEST	Kermit O. Hanson, *14034 24th Avenue, N.E., Seattle, Washington*
PHILADELPHIA	Kenneth M. Snader, *Federal Reserve Bank of Philadelphia, Philadelphia 1, Pennsylvania*
SACRAMENTO STATISTICAL ASSOCIATION	Marie Vida Ryan, *3551 12th Ave., Sacramento, California*
SAN FRANCISCO	Kenneth C. Christensen, *245 Market Street, San Francisco 6, California*
ST. LOUIS	Kurt Pohlen, *Catholic Hospital Association, 1402 South Grand Boulevard, St. Louis 4, Missouri*
WASHINGTON STATISTICAL SOCIETY	Samuel Weiss, *Bureau of Labor Statistics, Department of Labor, Room 2119, Washington 25, D. C.*

JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION

VOLUME 44 JUNE 1949 NUMBER 246

ARTICLES

The Current Status of State and Local Population Estimates in the Census Bureau . . . HENRY S. SHRYOCK, JR., AND NORMAN LAWRENCE 157

The Uses and Usefulness of Binomial Probability Paper . . . . . FREDERICK MOSTELLER AND JOHN W. TUKEY 174

Teaching Statistical Quality Control for Town and Gown . . . . . EDWIN G. OLDS AND LLOYD A. KNOWLER 213

The Use of Sampling in Great Britain . . . . . C. A. MOSER 231

Unemployment and Migration in the Depression (1930–1935) . . . . . RONALD FREEDMAN AND AMOS H. HAWLEY 260

Minimum X^2 and Maximum Likelihood Solution in Terms of a Linear Transform, with Particular Reference to Bio-Assay . . . . . JOSEPH BERKSON 273

Some Inadequacies of the Federal Censuses of Agriculture . . . . . RAYMOND J. JESSEN 279

The Edge Marking of Statistical Cards . . . . . A. M. LESTER 293

Conrad Alexander Verrijn Stuart (1865–1948) . WALTER F. WILLCOX 295

PROCEEDINGS OF THE 108TH ANNUAL MEETING

Minutes of the Annual Business Meeting 297

Report of the Board of Directors 300

Report of the Secretary 303

Report of the Nominating Committee 304

Report of the Committee on Fellows 304

Minutes of the Meeting of the Commission on Statistical Standards and Organization 305

Report of the Treasurer 307

Report of the Auditors 308

BOOK REVIEWS

ANDERSON, J. L. AND DOW, J. B., *Actuarial Statistics: Vol. II, Construction of Mortality and Other Tables* . . . . . T. N. E. GREVILLE 311

CHAPIN, F. STUART, *Experimental Designs in Sociological Research* . . . . . MARGARET JARMAN HAGOOD 312

CHARLIER, C. V. L., *Elements of Mathematical Statistics* Including *Table of Poisson's Function by L. V. Bortkiewicz* . . . BURTON H. CAMP 313

. . . . . ALEXANDER M. MOOD 314

DOUGLASS, RAYMOND D., AND ADAMS, DOUGLAS P., *Elements of Nomography* . . . . . JOSEPH ZUBIN 315

GALLUP, GEORGE, *A Guide to Public Opinion Polls, Second Edition*. . . . . ROBERT COBB MYERS 315

HALD, A., *The Decomposition of a Series of Observations Composed of a Trend, a Periodic Movement, and a Stochastic Variable* . . . . . D. B. DELURY AND BOYD HARSHBARGER 317

KENNEDY, CLIFFORD W., *Quality Control Methods* . . . . . SEBASTIAN B. LITTAUER 320
. . . . . CHARLES R. SCOTT, JR. 322

MATÉRN, BERTIL, *Metoder att Uppskatta Noggrannheten vid Linje- och Provytetaxering. [Methods of Estimating the Accuracy of Line and Sample Plot Surveys]* . . . . . T. W. ANDERSON 323

VON MISES, RICHARD, *Lecture Notes on Mathematical Theory of Probability and Statistics* . . . . . BENJAMIN EPSTEIN 326

SUMNER, W. L., *Statistics in School*. . . . . F. G. CORNELL 327

VAN UVEN, M. J., *Mathematical Treatment of the Results of Agricultural and Other Experiments, Second Edition*. . . . . G. A. BAKER 329

WINTNER, AUREL, *The Fourier Transforms of Probability Distributions* . . . . . J. WOLFOWITZ 330

WOLD, HERMAN, *Random Normal Deviates: 25,000 Items Compiled from Tract No. XXIV (M. G. Kendall and B. Babington Smith's Tables of Random Sampling Numbers)* . . . . . H. BURKE HORTON 331

ZEISEL, HANS, *Say It With Figures*. . . . . GREGOR SEBBA 332

A cumulative Index to Volumes 1–34, 1838–1939, and Annual Indexes thereafter, may be obtained from the Office of the Secretary of the AMERICAN STATISTICAL ASSOCIATION.

JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION

Number 246 JUNE 1949 *Volume 44*

THE CURRENT STATUS OF STATE AND LOCAL POPULATION ESTIMATES IN THE CENSUS BUREAU*

HENRY S. SHRYOCK, JR.,
AND
NORMAN LAWRENCE
Bureau of the Census

EARLY IN THIS CENTURY, the Bureau of the Census was already experimenting with postcensal State and local population estimates based on such symptomatic data as the number of public utility consumers, voting registrations, counts from city directories, and school census figures. Estimates were apparently made by the simple ratio method, and it was decided that they were not so satisfactory as those made by a simple mathematical method. For many years, virtually all the published estimates were linear projections of the last two census figures. Later, when the growth of birth and death registration areas made possible fairly accurate estimates of the national total population, an apportionment formula was used to prorate the postcensal national growth among the States and cities in accordance with their previous relative growth.

Although many estimates by these two methods were published, the Bureau recognized that the average errors were large enough to cast doubt on their usefulness. Sudden changes in population trends during the early years of the depression of the 'thirties made it obvious that these mathematical assumptions could no longer be used. In 1936 the

* Adapted from a paper read by the first-named author before a Customer Administrative School for Public Health Executives and Vital Statistics Registrars at the International Business Machine Corporation, Endicott, N. Y., July 2, 1947.

The authors wish to express their appreciation to Mr. Benjamin Greenberg, who assisted in the preparation of the population estimates given here, and to Dr. Joseph F. Daly, who directed their attention to the new tests of statistical significance in an unpublished thesis by John Edward Walsh of Princeton University.

Bureau published its first series of State estimates by what has been called the "migration and natural increase method." Here the components of population change, natural increase and net migration, were estimated separately, the migration component being derived from a comparison of actual with expected school enrollment. Current State estimates were discontinued after 1937 pending the results of the 1940 census. It had been planned to resume publication after 1940 using an improved migration and natural increase method, based on school data or perhaps some other symptomatic series; but the war changed the situation in several ways.

The program of current population estimates in the Census Bureau received a windfall in the registrations for war ration books in 1942 and 1943. Primarily from this source, several series of State and county estimates were constructed for these years.[1] These were the most comprehensive such postcensal estimates based on empirical data ever published. On the average, they were undoubtedly the most accurate also.[2] With a few exceptions, no population estimates for cities or counties have been published by the Bureau since those for 1943.

The valuable war ration book figures were not, however, an unmixed blessing to producers and consumers of population estimates. Their existence interrupted the experimentation with the above-mentioned symptomatic series, such as school data, that are available in peacetime as well as in war years. A further delaying influence was the prospect of a national sample survey of population in October, 1946, from which population totals for States and large cities would have been forthcoming. The Congress appropriated money for the preparatory work on this survey, but in the following fiscal year it failed to approve the necessary remaining funds. From individual sample surveys in the fall of 1946 and April, 1947, the Census Bureau obtained estimates of total population for selected large cities and metropolitan districts, however. These will be discussed later.

Meanwhile, the demand for State and local estimates was mounting. People were conscious of the sweeping changes that had occurred since

---

[1] U. S. Bureau of the Census. *Population.* "Estimates of the civilian population by counties: May 1, 1942." Series P-3, No. 33, February 25, 1943.

U. S. Bureau of the Census. *Population.* "Estimates of the civilian population by counties: March 1, 1943." Series P-3, No. 38. October 31, 1943.

U. S. Bureau of the Census. *Population.* "Estimated civilian population of metropolitan counties, by single counties: March 1, 1943, and May 1, 1942." Series P-3, No. 40. January 7, 1944.

U. S. Bureau of the Census. *Population—Special Reports.* "Estimated civilian population of the United States, by counties: November 1, 1943." Series P-44, No. 3. February 15, 1944.

[2] Hauser, Philip M., and Tepping, Benjamin J., "Evaluation of census wartime population estimates and of predictions of postwar population prospects for metropolitan areas." American Sociological Review 9 (5): 473–480. Oct., 1944.

the outbreak of the war, and the importance of measuring them. Vital statisticians, market analysts, directors of planning boards, and a host of others had mailed or telephoned their urgent requests. The time of our small estimating staff was taken up with a wide variety of national estimates, and State estimates were next on the list. Within the foreseeable future, we knew we could not make estimates for all the counties and middle-sized cities. Many State health departments just had to have such estimates, however; and the question of giving them technical advice was raised.

METHODS USED RECENTLY

In 1946, Dr. Hope T. Eldridge, then of the Population Division of the Census Bureau, spoke before the Council on Vital Records and Vital Statistics. She described a method for making postcensal county estimates on the basis of school data, vital statistics, and figures on the armed forces. Later she expanded this paper, and it was published as "Suggested procedures for estimating the current population of counties."[3] A second and more detailed method was also outlined in this article.

The second method, slightly modified, was used in making most of the published State estimates for 1946 and 1947.[4] Estimates by both methods were also computed for all the cities surveyed in the autumn of 1946 and the spring of 1947 in order to have independent checks on the inflated sample figures. The prime purpose of the present report is to present what has been learned about these methods and their pitfalls. The Census Bureau would also be greatly interested in hearing of the experience of other workers with these methods.

The methods described in Dr. Eldridge's article may be summarized briefly here. Detailed examples of each method are available on request to the Bureau of the Census. The simpler method, Method I, assumes that the difference between the percentage change in elementary school enrollment for the local area and the national percentage change in population of elementary school age is equal to the percentage change through net migration to or from the local area. The base date for a State, city, or county may be November, 1943, or any other date for

[3] Bureau of the Census. *Population—Special Reports*, Series P-47, No. 4, April 30, 1947.

[4] U. S. Bureau of the Census. *Current Population Reports. Population Estimates.* "Estimates of the population of the United States, by regions, divisions, and States: July 1, 1940 to 1947." Series P-25, No. 12. August 9, 1948. This release also contained revised estimates for July 1, 1940 to 1945. The earlier State estimates for July, 1944 and 1945, were to a considerable extent simply extrapolations of those for November, 1943. Subsequent data on natural increase and shifts between civilian and military populations had been utilized, but net interstate migration had been estimated from that for the period 1940 to 1943.

which the total population is available. Other figures needed to complete the estimates, namely, natural increase, persons away in the armed forces on the base and estimate dates, and members of the armed forces stationed in the county on the estimate date, are handled in a fairly direct manner.

Method II differs from Method I only in a more elaborate technique of estimating net migration from the school data. By this technique, net migration is measured on the basis of the difference between the population of elementary school age as estimated from the school data and the expected population of that age had there been no migration since the base date. The expected population is computed by applying survival rates from an appropriate life table to the population cohort at the base date that became the population of elementary school age at the estimate date. This "aging" process must also be applied to cohorts of births occurring after the base date when estimates are being made for the later postcensal years. To illustrate, the expected population 6 to 13 years old on April 1, 1947, assuming no migration, comprises survivors of births between April 1, 1940, and March 31, 1941, as well as survivors of children under exact age 7 on April 1, 1940. Since data on the population by age are required for the base date, it is usually necessary to go back beyond 1943 to the last census. The "actual" population of elementary school age with which the expected population is compared is computed from the school datum for the estimate date by means of an appropriate ratio based on past experience. Some assumption must be made about the relation of the rate of net migration of the school-age population and that of the population of all ages, the simplest assumption being that the rates are the same. This is a critical assumption and is discussed more fully below.

COMPARISON OF METHODS

Comparisons have been made between estimates made by these methods and population figures from other sources. These sources include the following: (1) Special censuses (complete counts) taken by the Census Bureau at the request and expense of local governments; (2) Censuses of Congested Production Areas taken in 1944 by the Bureau in selected metropolitan counties—these were sample censuses with a coefficient of variation of less than one per cent for the total population of each Congested Area; (3) State censuses; (4) Estimates made from the registration for war ration books.

Estimates made by Methods I and II have been compared with population standards of these sorts for 48 States, 25 cities, and 102

counties. Some cities and counties appear more than once in the list, for different dates. Some of the standards are complete counts. Other standards are subject to a fairly large amount of error; but, on the whole, they should be quite accurate. The population standards for States are the November, 1943, estimates based on War Ration Book Number 4 and published in Series P-44, No. 3. For the 25 cities, the standards were as follows: Ration Book estimate, 7; State census, 7; Census of Congested Production Areas in 1944, 9; and Census Bureau special census, 2. The standards for the 102 counties were distributed as follows: Ration Book figure, 43; State census, 42 (4 in Florida, 35 in Kansas, and 3 in Massachusetts); Census of Congested Production Areas, 16; and Census Bureau special censuses, 1. Similar population standards were available for many other cities and counties, but school data had not been collected for them. All figures are for the civilian population. The estimate dates range from 1943 to 1947.

The results of these two empirical methods may also be compared with those from two methods that do not use local postcensal data. The first of these latter is simple arithmetic projection. The total population of each area was projected on the basis of the 1930 to 1940 trend. This figure was converted into one for the civilian population by assuming that the area had the same proportion of its *de jure* population in the armed forces as the national population did at the same date. The second method is the apportionment method used by the Bureau of the Census in the early 'thirties. It assumes that the area had the same share of the national increase in the postcensal period as in the last intercensal period. Because the civilian population decreased over most of the period in which we are interested, the apportionment method could not be applied in the form just described. We modified the usual method by using the total population, which as a final step was converted into the area's civilian population, again by assuming the same proportionate loss to the armed forces by each area as was lost by the United States as a whole. (In the case of areas that had lost population between 1930 and 1940, it was assumed that there was no further decline in the *total* population after 1940. This total was reduced as just described to get an estimate of the postcensal civilian population.)

There is not space to present here the various estimates and their deviations from the population standard for each of the 175 States, cities, and counties. A summary of the deviations is given in Table 1, however, for each type of area. Happily, the extra work of the school-data methods of estimation appears to be decidedly worthwhile. In

TABLE 1

SUMMARY OF PERCENTAGE DEVIATIONS FROM POPULATION STANDARD OF ESTIMATES BY FOUR METHODS FOR SELECTED AREAS AND DATES
(See text for description of methods)

Item	Number of Areas	Method I	Method II	Arithmetic Projection	Apportionment
Average percentage deviation (disregarding sign)					
All areas	175	7.06	6.48	12.08	12.17
States	48	4.50	3.08	6.14	6.09
Cities	25	6.76	5.13	12.89	11.34
Counties	102	8.34	8.40	14.67	15.24
Root-mean-square percentage deviation					
All areas	175	9.43	9.29	16.89	16.75
States	48	5.32	4.29	7.44	7.29
Cities	25	7.75	7.35	14.54	12.96
Counties	102	11.16	11.23	20.31	20.37
Deviations of 10 per cent or above					
All areas	175	43	40	82	83
States	48	3	2	10	9
Cities	25	7	3	16	13
Counties	102	33	35	56	61
Positive deviations					
All areas	175	92	92	70	75
States	48	32	29	28	28
Cities	25	9	10	1	2
Counties	102	51	53	41	45

each type of area, the arithmetic mean of the percentage deviations disregarding sign is lower for both Method I and Method II than for either of the two mathematical methods, arithmetic projection and apportionment. (This fact does not gainsay that for some individual areas the percentage deviation was lower on the part of the mathematical than of the school-data method.) Average "errors" of 6 or 7 per cent over a period of 3 to 7 years by the better methods may not seem satisfactory for some purposes, but at least they compare favorably with the 12 per cent that would have resulted from the use of conventional methods.

It is of interest to investigate whether any of the differences between the average per cent deviations shown in Table 1 might reasonably be

attributed to chance variation. In computing the significance of the average percentage deviation for one method as compared with another, one should take account of the correlation between the deviations by the one method with those by the other. It is not surprising that there is a correlation between the deviations by Methods I and II and also between those by arithmetic projection and apportionment. (For all areas, there is a correlation coefficient of +.82 between the first two methods and of +.95 between the last two.) It is perhaps surprising, however, that there are also significant positive correlations between the deviations of a school-data method and those of a mathematical method, r ranging from +.36 to +.47. All correlations are positive within each type of area, States, cities, and counties. In other words, since the mathematical and empirical methods have no common assumption, a population that is difficult to estimate by one type of method also tends to be difficult by the other.

In testing the significance of the difference between the averages of two sets of deviations, we have in all cases allowed for the positive correlation by getting the difference between paired percentage deviations (disregarding sign) from two methods and testing whether the mean difference is significantly larger than zero. We thus obtain for all areas combined:

	t	*P*
Between Method I and Method II (7.06 vs 6.48)	1.98	.05
and Arithmetic projection (7.06 vs 12.08)	5.91	<.001
and Apportionment (7.06 vs 12.17)	6.24	<.001
Method II and Arithmetic projection (6.48 vs 12.08)	7.02	<.001
and Apportionment (6.48 vs 12.17)	7.15	<.001
Arithmetic projection and Apportionment (12.08 vs 12.17)	0.34	.7

Thus the superiority of either of the school-data methods over either arithmetic projection or apportionment is almost certainly not due to chance. The greater accuracy of Method II than of Method I may be deemed of border-line statistical significance as far as this test is concerned.

There may be some doubt about the propriety of combining all three types of areas in the above examination. Table 1 shows, however, that the populations of States, cities, and counties are alike in being more accurately estimated by a school-data method than by a mathematical method. For States, cities, and counties considered separately, the probability that these differences arose by chance alone is always less than .01 and usually less than .001. Method II is significantly more

accurate on the average than Method I only for States. Its lower average "error" for cities could have occurred by chance about 1 in 10 times, and for counties the average accuracy is about the same.

Two measures presented in Table 1 stress the larger percentage deviations from the population standard, whether positive or negative. The square root of the mean-squared percentage deviation when used as a criterion does not change our evaluation of the relative accuracy of the four methods very much. Percentage deviations of 10 per cent or more are next shown, and once again the school-data methods with their separate handling of migration and natural increase appear more accurate than the mathematical methods. By neither criterion, however, is Method II significantly more accurate than Method I.

The tests employed above, of course, depend for their validity on the assumption that we are dealing with functions of normally distributed variables. The failure of these tests to differentiate conclusively between the merits of Methods I and II might, therefore, be due to a failure of the normality assumption. To check on this point a test must be used in which the level of significance does not depend on the distribution law of the variables concerned. Recent developments in the theory of order statistics have provided such tests.[5] The application of such a test indicates that estimates prepared by Method II have an average absolute deviation that is definitely smaller than that for estimates prepared by Method I. When the estimates for States, cities, and counties combined are considered the difference proved significant at the .02 level. For States alone the difference proved significant at the .005 level, and for cities alone at the .05 level. For counties alone the difference was not significant. Accordingly, it appears that the additional work required in using Method II instead of Method I will often be justified.

For the cities and counties, the deviations of the estimates from the school-data methods are more evenly divided between positive and negative values than are those from the other methods. This situation is not true for the States. The higher rate of growth in the 'forties than in the 'thirties is particularly apparent in the nearly universally negative deviations for cities shown by the mathematical methods.

In the estimates made by Method II, the school datum was translated into an estimate of the actual population of elementary school age by applying the 1940 ratio of the two statistics. Since this ratio had been variable prior to 1940, it was supposed that the accuracy of

---

[5] John E. Walsh, "Some significance tests for the median which are valid under very general conditions." Unpublished doctoral dissertation, Princeton University Library, Princeton, N. J.

the final population estimates would be improved if the value of this ratio were extrapolated in some other way to the given postcensal date. We did this in two ways: (1) linear extrapolation of the trend in the ratio between 1930 and 1940; (2) logarithmic extrapolation. The latter assumes that the amount of change in the ratio has been progressively smaller. These two methods of extrapolation have been named Method IIa and Method IIb.

In Table 2, the final results (in terms of percentage deviation of the estimate from the population standard) from Methods IIa and IIb are compared with those from the original Method II, in which the ratio under discussion was held constant. A consistent series of school data was not available from 1930 to the estimate date for all areas, but the methods could be compared for a total of 132 States, cities, and counties.

TABLE 2

SUMMARY OF PERCENTAGE DEVIATIONS FROM POPULATION STANDARD OF ESTIMATES BY THREE VARIATIONS OF THE "MIGRATION AND NATURAL INCREASE" METHOD FOR SELECTED AREAS AND DATES

Item	Number of Areas	Method II	Method IIa	Method IIb
Average percentage deviation (disregarding sign)				
All areas	132	6.20	7.13	7.68
States	32	3.38	4.67	4.72
Cities	19	3.98	4.59	4.53
Counties	81	7.83	8.70	9.58

It may be readily observed that these refinements of Method II do not improve on the original. It would appear that there was either very little change in the ratio of elementary school enrollment to population of elementary school age after 1940 or else the change was of a quite different nature from that between 1930 and 1940. Until at least 1950, we may safely spare ourselves the extra work involved in these particular refinements of Method II. On the other hand, the extra work of Method II as compared with Method I seems to offer promise of more accurate population estimates.

A second source of data that may be used to assess the comparative accuracy of Methods I and II is available in the results of sample surveys taken by the Bureau of the Census in October and November, 1946, and in April, 1947. In 19 cities twice the coefficient of variation for the sample estimate of the civilian population of the city concerned

was about 5 per cent. Forty-two other surveys were based on smaller samples and are subject to still larger sampling errors. Moreover, the sample estimate for each of the 61 surveys was subjected to an upward adjustment of about 4 per cent on the assumption that this adjustment would make the results more comparable with those that would be obtained by a complete census. Evidence from national sample surveys indicates that an upward adjustment of about 4 per cent is needed to make the national Current Population Survey estimates comparable with those which would be obtained from a complete census. The case for applying the same adjustment to the results of local surveys, although convincing, was not altogether conclusive; and, in any event, the adjustment may well have been too high or too low for particular cities.[6] Consequently these sample estimates cannot be regarded as having the same status as the population standards referred to previously. We can merely examine the general consistency of the sample-based estimates with the result for each of the two synthetic methods. Such comparisons are presented in Table 3.

TABLE 3

SUMMARY OF PERCENTAGE DEVIATIONS FROM SAMPLE CENSUS FIGURES OF ESTIMATES BY METHODS I AND II: 61 SELECTED CITIES, 1946 AND 1947

Item	Method I	Method II
Average percentage deviation (disregarding sign)	7.6	4.9
Root-mean-square percentage deviation	9.3	6.3
Deviations of 10 per cent or above	15	6
Positive deviations	19	32
Negative deviations	42	29

Measured by both the average deviation and the root-mean-square percentage deviation, the more elaborate Method II gave results more consistent with the sample estimates than did Method I. The respective average deviations were 4.9 and 7.6 per cent; the respective root-mean-square percentage deviations were 6.3 and 9.3 per cent. These comparisons are based on 61 cases. When we pair the deviations (disregarding sign) from the sample estimate for Method I and Method II, compute the mean difference of the paired deviations, and examine by the t-test whether this mean difference (2.7 per cent) is significantly different from zero, we find that the probability that the deviations from the two methods could have come from the same universe is less than .001.

[6] If this adjustment had been as little as 1 per cent, the conclusions set forth below would not have been changed materially.

A very low probability was also obtained by the previously cited Walsh test.

Our preferred school-data estimates also come much closer than the simpler school-data estimates to an equal distribution of positive and negative deviations from the sample estimates. Method II has 32 positive and 29 negative deviations, whereas Method I has 19 and 42.

Thus the evidence seems to confirm our choice of Method II. Of course, some third method might be even better. Where the school-data estimate and the sample estimate for a city agree, we have some feeling of confidence in both since they are from entirely independent methods. Where the estimates disagree, either or both may be in error.

EVALUATION OF SCHOOL DATA

Inasmuch as both Methods I and II depend largely on school statistics for the measurement of net migration, serious errors of estimation can creep in at the very beginning of the estimating process unless the school data used are assembled with the main purpose always in mind. In either method school statistics are needed to approximate as closely as possible the number of children in a fixed age group. Some types of school data do so better than others. For example, statistics relating only to children of compulsory school age are more useful than those relating to all school children regardless of age since many of the latter go to school only by choice. Thus, data on enrollment by age are to be preferred over data on enrollment by grade for the purpose of making population estimates. If the former are not available, as is usually the case, grade enrollment figures are next to be preferred over other types of pupil statistics. But when such figures are used they should be restricted to the elementary grades, which comprise pupils most of whom are within the ages of compulsory school enrollment.

Two other types of school statistics are usually available. The first, average daily attendance, should be avoided on principle. It is obvious that an increase in average daily attendance may reflect no increase in enrollment or population but merely improved school administration or better weather during the school year. Over the long run, improved economic conditions or changes in parental attitudes conducive to more regular attendance at school may also be important factors. As to the second of these types of data, the school census, experience has taught us to proceed cautiously before using it to estimate population. It is no simple problem to enumerate the population of an area with reasonable

completeness even when trained enumerators do the job under well organized administrative controls. Yet, in most instances, school censuses are taken by school teachers, policemen, or unpaid volunteers, often without the geographic and other administrative controls needed to assure reasonably complete enumeration. As a result, school census data frequently vary in accuracy from year to year making it difficult to estimate the population in an age group by reference to the relationship between the school census and the United States census. Children outside the compulsory enrollment ages are especially likely to be underenumerated relative to a federal census. Frequently school census data are not tabulated by age so that the totals shown comprehend several age groups among which the completeness of enumeration varies greatly. Instances have even come to our attention of reports involving the publication of mere estimates improperly labeled as school census returns. In view of these difficulties, the school census has only rarely been used by the Census Bureau, and then only after careful investigation of the data.

A series that closely reflects the size of a fixed age group may nevertheless contain a defect from the standpoint of measuring net migration for a given area. One such situation occurs when the area for which the estimate is being prepared is not the same as the area for which the school statistics are compiled. School districts, it is true, are legally constituted areas, but they are not always the political areas for which population estimates are desired. Thus school statistics for the San Antonio Independent School District include some pupils who do not live in the city of San Antonio and exclude some who do live there. To use the local figures, a school-by-school tabulation was required to eliminate those not in the city. Another instance in which the area of jurisdiction of the school system is a complicating factor is found when a city has annexed one or more suburban areas since the last decennial census. The legal status of the annexed area changes at an instant in time, but the necessary administrative arrangements to consolidate the annexed school into the city school system may take some time to complete. It may be a year or more before the pupils in the annexed area are reflected in the statistics for the city.

Another stumbling block in the use of school statistics is a concealed inconsistency. Frequently a series of school data appears reasonable and consistent, but subsequent investigation reveals an administrative act which renders the figures not only useless but misleading. Thus, in one instance, a sudden increase in elementary enrollment proved to represent nothing more than a change from a seven- to an eight-grade

elementary school system. There are several different ways in which such a change-over might be made; each will leave its impression on the statistics. In the instance mentioned, the impact of the change-over was felt all at once, creating a suspicion on our part that could be checked. If the impact is spread over many years, it may not be noticed and the data may appear to indicate a gradual population increase when all that is involved is an administrative change in the organization of the school system. Similarly the opening of kindergartens or the unannounced introduction of kindergarten pupils into the enrollment statistics will produce a misleading increase in a school data series. (In general, kindergarten pupils should be excluded from the school statistics because attendance for them is not compulsory. For similar reasons, children less than 7 years old, or at least those less than 6, should be omitted when the data are classified by age.) Still another illustration of essentially the same situation is the omission from the school data series of pupils in special or ungraded classes. The omission might become a matter of consequence if the school system is in the process of expanding such programs and is steadily reclassifying its pupils and assigning them into these classes from the regular classes.

One factor about which statisticians have been apprehensive has usually proved to be unimportant. This is the presence of nonresidents in the local schools. Nonresidents, however, are seldom found in elementary classes because of the travel hazards for young children. Children from out of town are more frequently found in high schools; this fact is another reason for not using high school enrollment data.

The extent to which such circumstances as those mentioned can affect the population estimates may sometimes destroy whatever value inheres in the method of estimation. One of the concealed inconsistencies cited above resulted in an estimate of population 14 per cent greater than that obtained when the inconsistency was eliminated. Anyone using school statistics to estimate population must not only be aware of the requirements for data imposed by the estimating technique but must also be a statistical detective, going behind the data to the operating system from which they were derived, and understanding the changes taking place in educational administration as they affect school statistics.

In many areas, a substantial proportion of the population of elementary school age does not attend public school but receives its education at parochial or other private schools.

The importance of using enrollment figures that include private as well as public schools may be demonstrated by comparing estimates

based on total enrollment figures with those based on public enrollment. From the 61 cities of Table 3, 47 in which private enrollment constituted at least 10 per cent of total enrollment were selected. Two sets of estimates were then made by Method II. The results were compared with the population returned by the sample censuses already discussed.

TABLE 4

SUMMARY OF PERCENTAGE DEVIATIONS FROM SAMPLE CENSUS FIGURE OF ESTIMATES BY METHOD II, USING TOTAL ENROLLMENT AND PUBLIC ENROLLMENT: 47 SELECTED CITIES, 1946 AND 1947

Item	Using Total Enrollment	Using Public Enrollment
Average percentage deviation (disregarding sign)	4.7	8.9
Root-mean-square percentage deviation	5.8	10.1
Deviations 10 per cent or above	3	19
Positive deviations	23	8
Negative deviations	24	39

When parochial, and sometimes other private, enrollment is taken into account, the population estimates are significantly more consistent with estimates obtained by sample censuses. The average absolute percentage deviation for the estimates based on both public and private enrollment (4.7 per cent) was significantly smaller ($P<.001$) than the average based on public enrollment alone (8.9 per cent) by either the t-test or the test devised by Walsh. The improvement is so pronounced that a strong case is made for collecting school data covering parochial and other private schools. The fact that Method II using public school data tends to give low estimates reflects the shift in recent years away from public schools and to parochial schools.

OTHER SOURCES OF ERROR IN METHOD II

Method II almost always requires the use of census statistics on children under 5 years old. In the census of 1940, as in most other censuses here and abroad, it is known that many children of this age were not enumerated. The census figures therefore have to be corrected for underenumeration of young children before the number of expected survivors is computed. Failure to do so would result in understatement of the expected population of school age and a commensurate decrease in the estimated net migration.

There are published measures of underenumeration for States, urban and rural; but for other areas one must select the measure that seems most appropriate from a priori considerations. There are, of course, intercounty variations in underenumeration within a State just as

there are in the completeness of birth registration. The variation is probably greatest in the States where underenumeration is greatest. For 1940 the estimated completeness of enumeration ranged from 79.9 per cent for South Carolina nonwhites to 98.0 per cent for Nebraska and South Dakota whites.[7]

If one is dealing with areas smaller than States, it usually will be necessary to distribute some age groups by single years of age since the single-year data were not published. This step can be performed fairly well on the basis of expected survivors of resident births during the appropriate years preceding the census. Even in areas with heavy migration, only minor errors should be introduced by this assumption.

Another point at which relatively minor errors may occur is in "aging" the corrected population distributed by single years to the estimate date.

A life table for the given area covering the given time usually will not be available. Hence the most appropriate life table must be selected. In some cases, it may seem desirable to adapt an existing life table. Recent life tables published by the federal government cover: (a) regions, urban and rural, by color and sex, for 1939;[8] (b) geographic divisions, by color and sex, for 1930 to 1939;[9] and (c) the white population of States, by sex, for 1939 to 1941.[10]

Although there is a fair amount of variation among areas in infant and child mortality rates, there is not much variation percentagewise in their complement, survival rates. For instance, according to 1929–1931 life tables for white males, by States, the proportion of the cohort under 7 years old surviving to be 7 to 13 years old varies only from .976 for Colorado to .986 for South Dakota. Nonetheless, it would be best to construct a life table for the area concerned, if it is large enough to have stable mortality rates. Actual deaths should not be subtracted from the population of the cohort at the base date since they include those of in-migrants and exclude those of out-migrants. It will be recalled that we are estimating the expected survivors assuming no migration.

Still another source of error arises from the fact that the ratio between the population of elementary school age and elementary school enrollment for any area is not constant over time. For States the

---

[7] U. S. Bureau of the Census. *Population*. "Differential fertility: 1940 and 1910. Standardized fertility and reproduction rates." p. 33. Washington, Government Printing Office, 1944.

[8] U. S. Bureau of the Census. "U. S. Abridged Life Tables, 1939, Urban and Rural, by Regions, Color, and Sex." June 23, 1943.

[9] U. S. Bureau of the Census. "U. S. Abridged Life Tables, 1930–39 (Preliminary), by Geographic Divisions, Color, and Sex." April 30, 1943.

[10] National Office of Vital Statistics. "State and Regional Life Tables, 1939–41." Washington, Government Printing Office. 1948.

average change in the ratio of the population 7 to 13 years old to public elementary school enrollment was +0.038 between 1930 and 1940. Disregarding sign, the average change was 0.047. The range was from −0.052 to +0.141. It has been shown in Table 2, however, that during the 'forties more accurate population estimates are obtained on the average by assuming that the 1940 ratio did not change than by extrapolating the 1930–1940 trend in this ratio.

We come now to what is probably the major assumption in our method. This is that the net rate of change through migration for elementary school-age children is equal to the net migration rate for the population of all ages. Migration is quite selective of certain age groups, of course; but there is some evidence that children of elementary school age, who move with their parents, are of intermediate mobility between young adults and the aged. Insofar as child migration is near the average rate, it may be used to represent the rate for the whole population.

Past experience in the pattern of net migration rates by age may be studied. From 1935 to 1940, the net interstate migration of children 5 to 13 years old was in the same direction as that of the total population in all except four States and the District of Columbia. Among the 44 States where net migration was in the same direction, the rate for children 5 to 13 years old averaged 91 per cent of that for all ages. The range was from 40 per cent to 167 per cent, with a standard deviation of 35 percentage points.

Among the 90 cities of 100,000 or more in both 1940 and 1930, the net migration of children 5 to 13 years of age was in the same direction as that of the total population in all except 5 cases. In these 85, the rate for the children averaged 145 per cent of that for all ages. The range was from 3 per cent to 400 per cent, the standard deviation being 69 percentage points.

From these facts, it can be appreciated that there is a great deal of variation in the relationship between the child-migration rate and the rate for the population of all ages. It is thus easy to make a large error in estimating the total net migration from the migration at the age of school enrollment. We can, of course, use the ratio observed in the particular State or city with which we are dealing. We do not know, however, how representative the ratio for 1935 to 1940 is for subsequent periods. There are still other disturbing factors. One is that in making current postcensal estimates we have to deal with periods ranging from one to nine years instead of just five years. Another is that the age group chosen to represent school enrollment may not be 5 to 13; for

instance it may be 7 to 14. Late in the decade, the school-age cohort includes some children born since the last census. This part of the cohort has had a shorter period of exposure to "risk" of migration than have persons of older ages. Fortunately, there is never a very large difference in the average length of exposure between the whole school-age cohort and the population of all ages.

A great deal of investigation remains to be done on the age selectivity of net migration for different areas over various periods of time. So far we have relatively few statistics on net migration by age. If we have taken care of the other sources of error in the method so that the migration assumption is the only important item of guesswork remaining, we can do some experimentation by trial and error. We can see what migration assumptions produce the most accurate population estimates, using a complete census or other reliable figure as the standard.

The foregoing discussion of sources of error was concerned with the estimation of net migration. The estimation of the actual natural increase is beset with relatively minor pitfalls. The available figures on under-registration of births are subject to some error. It is usually assumed that deaths are almost completely registered. There may also be some inaccuracy in the allocation of births and deaths to the correct area of usual residence.

The Census Bureau's State estimates, which are prepared essentially by Method II, are adjusted to total to a national estimate, which we consider to be quite comparable to a decennial enumeration. Similarly, county estimates made locally may be scaled up or down to a State total. Such controls reduce the chance of very large errors. Whenever censuses or better estimates are available for particular local areas, they should, of course, be substituted. For example, the method we have described proved inappropriate for Washington, D. C., because of the unusual character of migration to this city. As an experiment, this Bureau used a wide variety of statistics in preparing the estimates recently published.[11] We feel that although our school-data method is promising enough to warrant extensive experimentation and polishing, concurrent exploration of other sources and methods should be conducted.

[11] U. S. Bureau of the Census. *Population—Special Reports.* "Estimated population of the Washington, D. C., Metropolitan Counties: 1940 to 1946." Series P-47, No. 5. May 14, 1947.

THE USES AND USEFULNESS OF BINOMIAL PROBABILITY PAPER*

FREDERICK MOSTELLER
AND
JOHN W. TUKEY
Harvard University and Princeton University

This article describes certain uses of Binomial Probability Paper. This graph paper was designed to facilitate the employment of R. A. Fisher's inverse sine transformation for proportions. The transformation itself is designed to adjust binomially distributed data so that the variance will not depend on the true value of the proportion p, but only on the sample size n. In addition, binomial data so transformed more closely approximate normality than the raw data.

The usefulness of plotting binomial data in rectangular coordinates, using a square-root scale for the number observed in each category, was first pointed out by Fisher and Mather [10]. The graph paper under discussion[1] is specially ruled to make this mode of plotting both simple and rapid. A graduated quadrant makes the angular transformation ($p = \cos^2 \phi$ or $p = \sin^2 \phi$) easily available at the same time. Most tests of counted data can be made quickly, easily and with what is usually adequate accuracy with this paper. Some 22 examples are given.

PART I—GENERAL

INTRODUCTION

IN ANALYZING data which has been counted rather than measured the biologist, opinion pollster, market analyst, geologist, physicist, or statistician frequently uses as a model the binomial distribution, its limiting case the Poisson distribution, or some of their generalizations. For many purposes graphical accuracy is sufficient. The speed of graphical processes, and more especially the advantages of visual presentation in pointing out facts or clues which might otherwise be overlooked, make graphical analysis very valuable. A special type of graph paper has been designed and is now available for such analysis.

The examples below show how such paper can be used in various computations—clearly there are many parallel cases in other fields of analysis.

* Prepared in connection with research sponsored by the Office of Naval Research.

[1] Available from the Codex Book Company, 74 Broadway, Norwood, Mass. No. 31,298 on thin paper, No. 32,298 on thick paper.

The basic idea involved is due to R. A. Fisher, who, many years ago, observed that the transformation

$$\cos^2\phi_i = \frac{n_i}{n}$$

transformed the multinomial distribution with observed numbers $n_1, n_2, \cdots, n_k$ into direction angles $\phi_1, \phi_2, \cdots, \phi_k$ which were nearly normally distributed with individual variances nearly $1/4n$ (when the angles are measured in radians). Thus the point at a distance $\sqrt{n}$ from the origin and in the direction given by $\phi_1, \phi_2, \cdots, \phi_k$ is distributed on the $(k-1)$ dimensional sphere nearly normally, and with variance nearly $\frac{1}{4}$ *independent of n and the true fractions* $p_1, p_2, \cdots, p_k$ *of the different classes in the population.* The rectangular coordinates of this point are $\sqrt{n_1}, \sqrt{n_2}, \cdots, \sqrt{n_k}$.

In 1943 Fisher and Mather [10] applied this principle to the graphical plotting of the square roots of the observed numbers of *Lythrum salicaria* styles of various types on ordinary graph paper. After reading this article, the present authors were convinced that special graph paper, designed to facilitate this process would be worthwhile and designed the paper [12] mentioned above—whose use forms the subject of this article.

Binomial Probability Paper is graduated with a square-root scale on both axes 1(1)20(2)100(5)400(10)600 and 1(1)20(2)100(5)300. (The notation $A(a)B(b)C$ means: "From A to B by intervals of size a and thence from B to C by intervals of size b".) It serves to treat classification into two categories directly. Cases with three categories would require three-dimensional graphing for direct treatment, but a device not explained in this paper enables one to treat any number of categories on this paper.

PLOTTING

The binomial distribution

The classical example of the binomial distribution is the distribution of the number of heads in n independent throws of a coin, true (50-50), or biased (for example 40-60 or 57-43). Two assumptions require emphasis here (i) the throws are independent and do not influence one another, (ii) the probability of a head on each throw is the same. Replace "heads" and "tails" by "successes" and "failures," or by "regular females" and "exceptional females," or by "alcoholic" and "non-alcoholic," or by any classification into two categories and these two

requirements, suitably modified, are the requisite conditions that the observed numbers in the categories will be distributed binomially.

We must distinguish (i) the true relative probability of two categories, which we will often refer to as the *split* and (ii) the observed numbers in the two categories, which we will often refer to as the *paired count.* In describing a split we do not bother with normalization —the split of a "true" coin is 50-50 or 1-1 or 127-127. In describing a paired count we always give the actual numbers counted.

Binomial plotting

A *split* is plotted as a straight line through the origin—it passes through all the points whose coordinates represent it. Thus in Figure 1 the lowest line corresponds to the 80-20 split. In addition to going through (80, 20) this line goes through the points (8, 2), (16, 4), (20, 5), (120, 30) and so on—the corresponding splits are all the same. The 77-23 split is also called the 23% line, and the (100-X)-X split is also called the X% line.

A *paired count* is plotted as a triangle whose sides extend for one scale unit in the positive directions of the axes from its right angle, which has the paired count for its coordinates. Thus (1, 3) is plotted as the triangle with vertices (1, 3), (2, 3) and (1, 4). The paired counts (1, 3), (14, 1), (155, 4), (1, 110), (125, 150) are plotted in Figure 1. (The modification of the classical angular transformation thus induced is discussed in Part V (page 206).)

If any coordinate exceeds 100, it is usually satisfactory to plot a line segment (vertical in the case of (155, 4) and horizontal in the case of (1, 110)). If both coordinates exceed 100, it is usually satisfactory to plot a point.

The middle of a paired count is the center of the hypotenuse, which degenerates into the center of the line segment.

Percentages

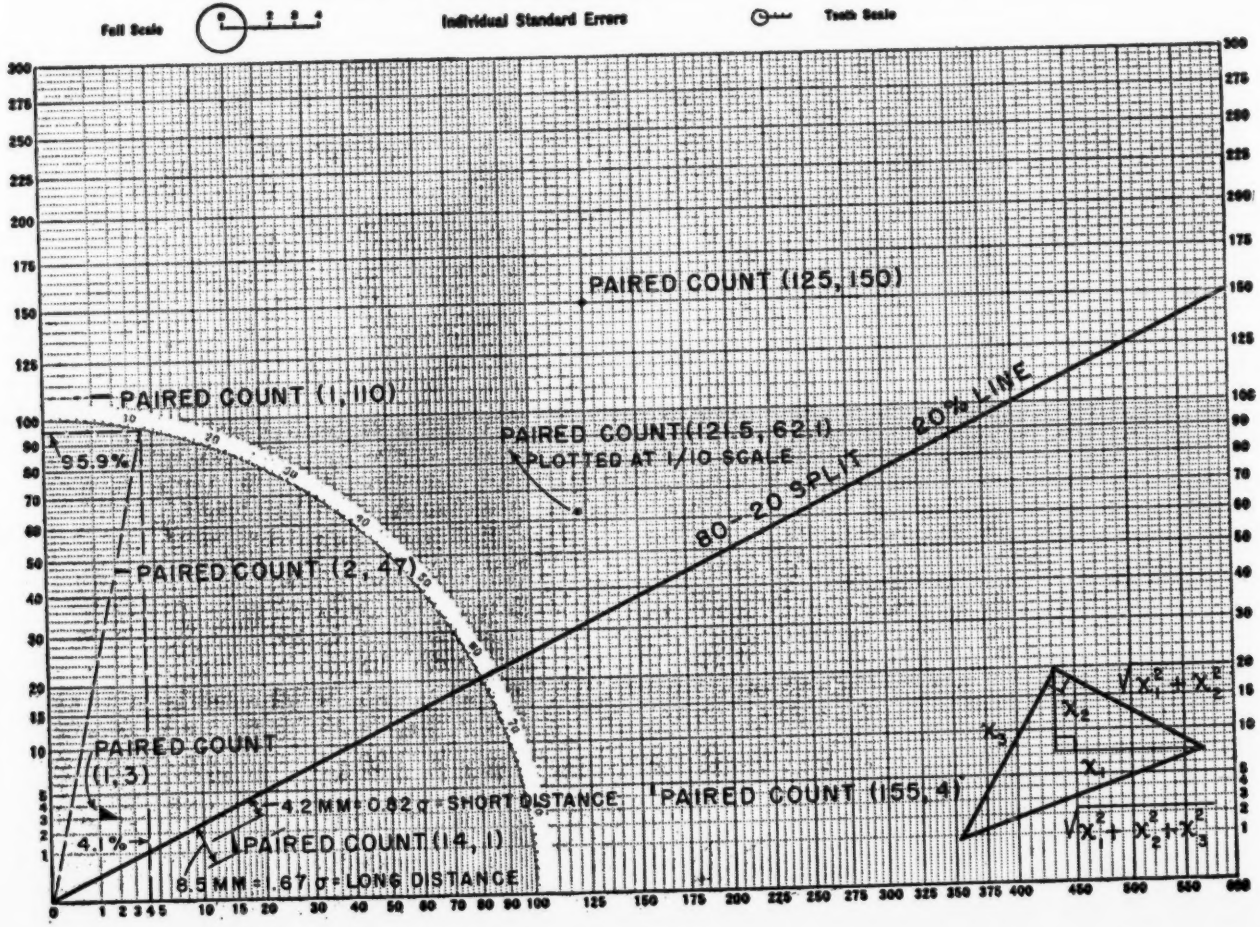
1. *To convert a split into a theoretical percentage read the coordinates of its intersection with the quarter circle* (paying no attention to the angular scale).

2. *To convert a paired count into an observed percentage, draw the split through its lower left corner and read the coordinates where the split crosses the quarter circle.* (It is convenient to consider this an auxiliary line and draw it dashed). Thus (2, 47) corresponds to percentages of 4.1 and 95.9 (Figure 1).

FIGURE 1
SHOWING PLOTTING (PAIRED COUNTS, SPLITS, 1/10 SCALE), MEASUREMENT OF DISTANCES (SHORT AND LONG), CALCULATION OF PERCENTAGES, AND CRAB ADDITION.

FIGURE 1

SHOWING PLOTTING (PAIRED COUNTS, SPLITS, 1/10 SCALE), MEASUREMENT OF DISTANCES (SHORT AND LONG), CALCULATION OF PERCENTAGES, AND CRAB ADDITION.



Measurements of deviations

A right triangle has three corners, but we shall only be concerned here with the two acute-angled ones. There are two distances from a line to these corners of a triangle. We call these the *short distance*, and the *long distance*. Thus, in Figure 1, the distances from the 80–20 split to the triangle representing (14,1) are 4.2 mm. and 8.5 mm. The distance to the middle of the hypotenuse, which is the average of the other two, is the *middle distance*, here 6.4 mm.

In some cases there are definite reasons, theoretical or empirical, to prefer the short distance, the long distance or a combination of the two. In the remaining cases it seems plausible to use the middle distance, though investigation may prove otherwise. In the examples below we have followed this plausible procedure.

Such distances are usually to be interpreted as nearly normal deviates. The scale in the upper left marked "Full scale—Individual Standard Errors" allows these distances to be read as multiples of the standard deviation.

For most purposes, however, a millimeter scale is more convenient. On the paper *one standard deviation is almost exactly 5 mm.* (if the paper has not shrunk or stretched, it is 5.080 mm.). Thus the short, middle and long distances of 4.2, 6.4 and 8.5 mm., from the example above, are 0.82, 1.24 and 1.67 standard deviations. A short table of the normal distribution is given in Part VI, both in millimeters and in standard deviations. For many purposes it is only necessary to remember that one deviation in three should be outside −5 mm. to 5 mm. and one in twenty outside −10 mm. to 10 mm. purely due to sampling fluctuations.

The reader who is interested in the simpler applications should now go to Part II (page 182). The reader who is interested in a specific application is referred to Part VI (page 209) which provides an index and outline.

REFINEMENTS IN PLOTTING

Plotting larger counts

Paired counts up to (600, 300) can be plotted directly. If there are more observations, but less than (6,000, 3,000), the paired count may be plotted at 1/10 scale, thus (1,215, 621) would be plotted as (121.5, 62.1) as in Figure 1. In this case

(1) the left-hand upper scale should be replaced by the right-hand upper scale in converting distances into multiples of the standard deviation;

(2) distances in millimeters should be multiplied by 3.16 before using Tables 1 and 2;
(3) the triangle would extend from (121.5, 62.1) to (121.6, 62.1) and (121.5, 62.2) and would be completely indistinguishable from a point.

Very large counts

If significance tests are required for still larger samples, graphical accuracy is insufficient, and arithmetical methods are advised. A word to the wise is in order here, however. Almost never does it make sense to use exact binomial significance tests on such data—for the inevitable small deviations from the mathematical model of independence and constant split have piled up to such an extent that the binomial variability is deeply buried and unnoticeable. Graphical treatment of such large samples may still be worthwhile because it brings the results more vividly to the eye.

Plotting unsymmetrical counts

Occasionally the user may have data involving large counts of one kind and small counts of the other, where the total count may not be fixed. Such cases may be treated by choosing a suitable divisor, and using the divisor only for the category with the large counts. If this is done, the distance from paired count to split should be measured at right angles to the contracted axis and *not* at right angles to the split. Each triangle reduces to a segment extending one scale unit away from the divided axis. For good accuracy, it is necessary that each small count be less than 5% of the corresponding total count.

Poisson subcase

The Poisson distribution yields the subcase of this case where the total number is arbitrarily large and *fixed*. Binomial probability paper may still be used without difficulty, plotting the various observations on any chosen, fixed vertical line, thus using an unknown horizontal divisor.

Crab addition

In order to sum the squares of deviations in a chi-square, we may proceed either by measuring, squaring and adding, or by crab addition, using the theorem of Pythagoras. Crab addition is easily done by using a piece of tracing paper as follows:

i) Mark a small circle as the origin (0) on the tracing paper and set

it on the theoretical line at the point of intersection with the perpendicular dropped from the first sample point;

ii) Mark a point (1) on the tracing paper at the sample point—the distance between the tracing paper origin and the first sample point is the square root of the contribution of the first sample point to χ^2, call it χ_1^2.

iii) Line up the origin (0) and the point (1) on the tracing paper on the theoretical line with the point (1) at the base of the perpendicular from the second sample point to the theoretical line, and plot a point (2) on the tracing paper immediately over the second sample point. The distance from point (1) to point (2) is χ_2 the square root of the contribution of the second sample point to χ^2, and the distance from the origin (0) to point (2) is $\sqrt{\chi_1^2+\chi_2^2}$. We continue in this manner producing a picture like that shown in the lower right-hand corner of Figure 1.

Use for chi-square

If the individual segments submitted to crab addition were *middle distances*, then the final crab sum may be (geometrically) doubled in length and read on the marginal scale to obtain a χ^2 value. (Multiply this answer by another factor of 10 if the points were plotted to 1/10 scale.)

If the individual segments submitted to crab addition were *ranges* then the final crab sum may be read on the marginal scale and then (numerically) doubled in value to obtain a χ^2 value (see Ex. 24). (Multiply by another factor of 10 if the points were plotted to 1/10 scale.)

SIGNIFICANCE LEVELS AND SIGNIFICANCE ZONES

The continuous case

When a value of a continuously distributed statistic, such as Student's "t," is computed from a sample, there is no difficulty (provided the distribution is well tabled) in finding and stating its level of significance. Thus a t of -0.98 on 3 degrees of freedom lies at (i) the lower 20% point, (ii) the upper 80% point, (iii) a (balanced) two-sided 40% point. These statements mean that, *if* the simple situation *against* which we are assessing the evidence really holds, and similar experiments are repeated, (i) an algebraically smaller value (that is, <-0.98) of t will occur in 20% of the cases, (ii) an algebraically larger value (that is >-0.98) will occur in 80% of the cases, (iii) a value further from the center (that is, $|t|>0.98$) will occur in 40% of the cases. The two-sided statement is apparently simple to make in the case of t, since

the distribution concerned is symmetrical and continuous. In asymmetrical cases, we will follow the path of least resistance and calculate the two-sided significance level as twice the (smaller) one-sided significance level, just as 40% = 2(20%). (Some statisticians might disagree.) The choice of one among the three significance levels in a practical situation will depend on the alternatives considered for the situation against which the evidence is being assessed.

The discrete case

In the binomial situation, and in the many other cases where the result obtained proceeds in definite steps, the situation is a little more complex. Consider the case of a sample of 4 from a 2–1 split. The probabilities of the various possible outcomes are given below to 3 decimal places.

Outcome:	(0, 4)	(1, 3)	(2, 2)	(3, 1)	(4, 0)
Frequency:	.012	.099	.296	.395	.198
Cumulated from left to right:	.012	.111	.407	.802	1.000
Cumulated from right to left:	1.000	.988	.889	.593	.198

What significance level shall we assign to (1, 3) in this case ? The conventional answers are that it lies at: (i) the lower 11.1% point, (ii) the upper 98.8% point, (iii) the two-sided 22.2% point. These statements mean that, if the situation were binomial with a 2–1 split, (i) one of the outcomes (0, 4) or (1, 3) occurs in 11.1% of all cases, (ii) one of the outcomes (1, 3), (2, 2), (3, 1), (4, 0) occurs in 98.8% of all cases, (iii) it is reasonable to act as though an outcome deviating from expectation *as much or more* than (1, 3) in one direction or the other occurs in 22.2% of all cases.

An alternative approach, and one which supplies more information, is to attach to such a result as (1, 3) not a single significance level, but a *significance zone*. Thus the lower significance zone for (1, 3) is "11.1% to 1.2%" and the two-sided significance zone is 22.2% to 2.4% (where we have adopted the convention of doubling in passing from one-sided to two-sided significance zones.)

The statement "(1, 3) is at the lower (11.1%, 1.2%) zone" means that, if the simple binomial situation with split 2–1 holds, then (0, 4),

which is the only outcome further from expectation than (1, 3) *in the same direction*, will occur in 1.2% of all cases, while (0.4) and (1, 3) *together* will occur in 11.1% of all cases.

Interpretation

The working statistician, we believe, would almost always react differently to the statement—the outcome is at the lower (12%, 0.5%) zone—than to the statement—the outcome is at the lower (12%, 11.5%) zone. The (12%, 0.5%) zone indicates that we are not sure that strength of the evidence has reached the conventional 5% point, but it is possible that this is the case. The (12%, 11.5%) zone indicates that the strength of the evidence has certainly not reached the conventional 5% point. This distinction is absent from the customary procedure of assessing just a significance level.

In some cases, auxiliary experimental data could be used to interpolate between the ends of the significance zone.

As we shall see, binomial probability paper can be used to obtain both the customary significance level and the significance zone. If we are to determine the two percentages needed to fix a significance zone, we shall need to make two measurements on the paper, so it is not surprising that we will want to plot more than a single point to represent a paired count.

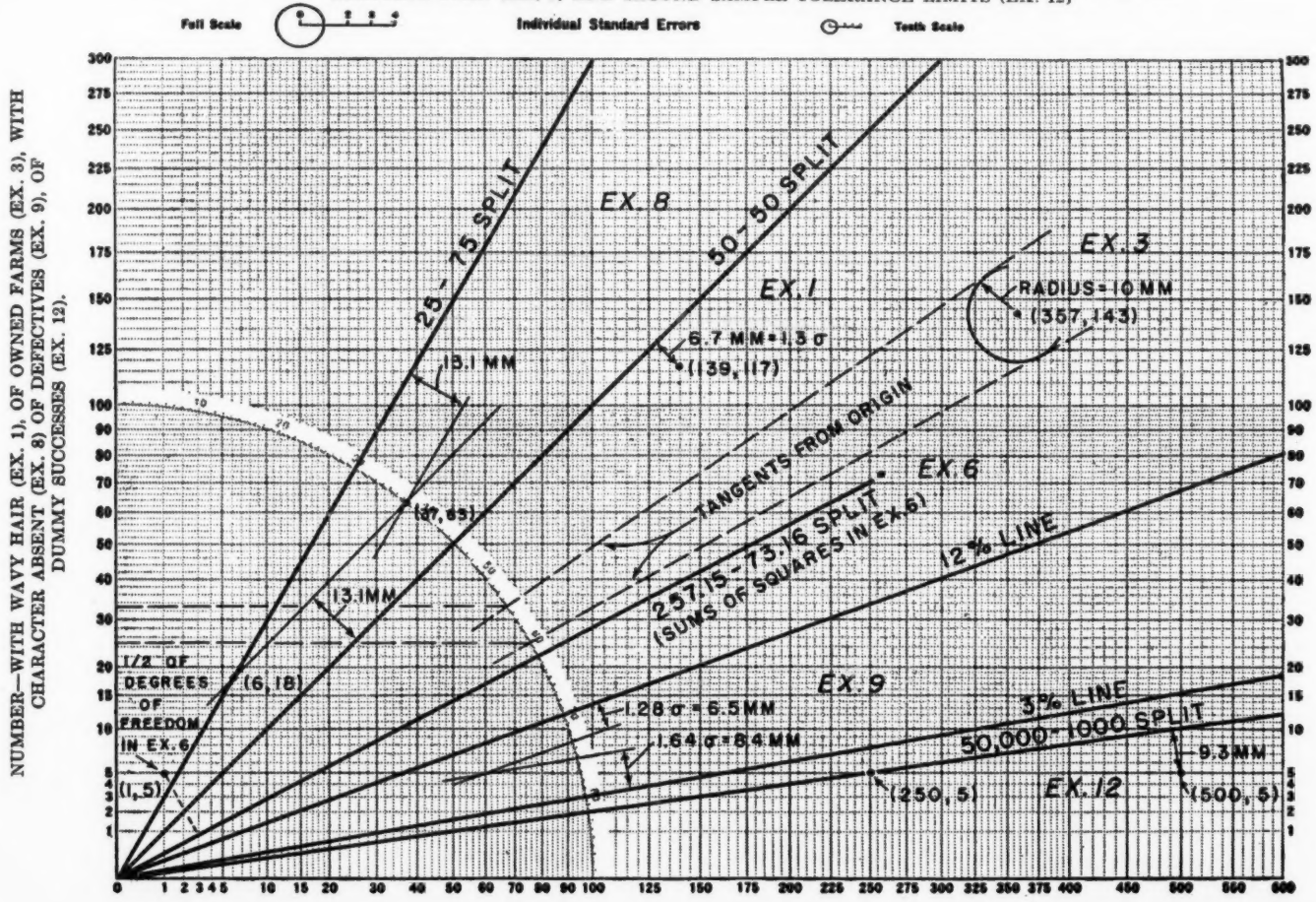
PART II—PLOTTING ONE OBSERVED QUANTITY

A SINGLE BINOMIAL SAMPLE

Given a sample, sorted into two categories, the numbers in the categories form a paired count, which is plotted as a triangle. If a probability or a population proportion, p, is assigned for the second category, then this is plotted as the q-p split, where $q=1-p$. The approximate significance level corresponds to the short distance from the split to the triangle, and may be obtained from Table 2 (page 211). If it is desired to use the significance zone, then both short and long distances should be measured and the probabilities obtained from Table 2. These general principles will suffice to attack the problems posed by the first four examples.

Comparing an observed proportion with a theoretical proportion

Example 1 (See Figure 2). Fisher and Mather [11] have described a genetical experiment in which the individuals in 32 litters of 8 mice were observed for the characteristic of straight versus wavy hair. Under the conditions of the experiment, the Mendelian theory



NUMBER—WITH WAVY HAIR (EX. 1), OF OWNED FARMS (EX. 3), WITH CHARACTER ABSENT (EX. 8) OF DEFECTIVES (EX. 9), OF DUMMY SUCCESSES (EX. 12).

NUMBER—WITH STRAIGHT HAIR (EX. 1), OF TENANT FARMS (EX. 3), WITH CHARACTER PRESENT (EX. 8), OF NONDEFECTIVES (EX. 9), OF DUMMY FAILURES (EX. 12).

predicted that half the mice would have straight hair. It was observed that $n_1 = 139$ had straight hair, $n - n_1 = 117$ had wavy hair. Could such a discrepancy from the 128-128 split have arisen by chance?

The observed paired count is plotted as the triangle (139, 117), (139, 118), (140, 117) and is just distinguishable from a point. The theoretical proportion is plotted as the 128-128 split (the 50% line). The short distance is about 6.7 mm. $= 1.3\sigma$. Since deviations from simple Mendelian genetics might reasonably occur in either direction, the two-sided significance level is appropriate, and from Table 2 this is found to be about 20%. Thus this test would not lead us to reject simple Mendelian genetics.

Since the two-sided 5% distance is 10 millimeters, which is worth remembering, Table 2 would not have been needed in routine testing, since the result "not significant at the 5% point" would have been enough.

We may observe the experimental percentage of straight-haired mice by drawing a horizontal line to the vertical axis from the intersection of the degrees quarter-circle with the line from the observed point to the origin. This gives about 54.5 per cent as compared with an actual value of 54.3 per cent.

If the significance zone were desired, then both the short distance of 6.7 mm. and the long distance of 7.5 mm. would have been measured. The corresponding two-sided significance interval is (20%, 14%).

Critique of Example 1. The experimental conditions seem to have been exactly suited for a binomial test. The accuracy of the graphical method is clearly adequate.

The sign test

The classical case of the sign test is the comparison of two materials or treatments, in pairs, where the observations in each pair are comparable except for the materials or treatments being tested. The sign test is a special case of the comparison of theoretical and observed proportions, where the theoretical proportion is always 50% and has been thoroughly discussed in this Journal by Dixon and Mood [4, 1946].

Example 2 (See Figure 3). Dixon and Mood cite the yields of two lines of hybrid corn where 6, 8, 2, 4, 3, 3, and 2 pairs of plots were available from 7 experiments. In 7 out of the 28 pairs line *A* yielded higher.

If a significance zone is wanted, then both the short distance of 12.9 mm. and the long distance of 14.7 mm. are measured. The resulting two-sided significance zone is (1.2%, 0.3%).

Comment on Example 2. The tables of Dixon and Mood correctly state that (7, 21) is not at the 1% level. Some statisticians, however, in part because of the general and unprecise considerations which lead to the choice of 1%, may use the extra information in the significance zone and decide that they would rather work at the 1.3% level than the 0.4% level (the precise values are 1.254% and 0.372%). They would then treat (7, 21) as significant at a level of approximately 1%.

Extension of the sign test

As presented above, the sign test applies to the hypothesis of equality in paired experiments. It can be easily extended to cover (i) the hypothesis of a constant additive difference, (ii) the hypothesis of a constant percentage difference, or (iii) the hypothesis of a certain population median by constructing dummy observations. Suppose that five experiments have produced the following numbers:

Set	1	2	3	4	5
Condition A	57	53	49	56	51
Condition B	26	31	24	28	31

There is clearly no need to test for equality. A test of the hypothesis that condition *A* runs 30 units above condition *B* may be made by adding 30 to each observation under condition *B*, which yields 2 positive and 3 negative differences. A test of the hypothesis that condition *A* gives numbers double those of condition *B* may be made by doubling the condition *B* numbers, which yields 2 positive, 1 zero, and 2 negative differences (the corresponding paired count is (2, 2)!). A test of the hypothesis that the median number in condition *A* is 57, may be made by replacing the results in condition *B* by an imaginary experiment always giving 57, which yields 1 positive, 1 zero, and 3 negative differences (the corresponding paired count is (1, 3)!).

Confidence limits for population proportions

If we have a sample divided into two categories, we may wish to set confidence limits on the percentage of the population in each category. This is accomplished by plotting the paired count which was observed and then constructing two splits whose short distances from this triangle correspond to the two-sided level (of confidence) required. Thus if 95% confidence is required, short distances of 10 mm. will be used.

The coordinates of the intersections of the two splits with the quarter-circle give the confidence limits for percentages.

Example 3 (See Figure 2). A random sample of 500 farms from a certain region yields the information that 143 of the farmers own outright the farms they work. Set 95 per cent confidence limits on the per cent of farms in this region wholly owned by the farmers.

We plot the triangle (357, 143), (358, 143), (357, 144), which can, in this case, be approximated by a point. We draw arcs of circles of radius 10 mm. about its extreme vertices, and draw the tangent splits. We get as our estimate of the percentage of farmer-whole owners 28.5 (compared to 28.6 computed) and as our 95% confidence limits, 25 per cent and 33 per cent.

Critique of Example 3. If the sample is truly random the method is sound. If stratified sampling is correctly done, these confidence limits will be unnecessarily large by an amount which depends on the effectiveness of the strata chosen, for the particular question at hand. The increase in efficiency by stratification is often so small that these limits are a wise choice.

Confidence limits for the population median

By combining the ideas of the sign test and the last example, we may obtain confidence limits for the population median. If x_0 is the population median, and is used to divide samples into paired counts, then certain extreme paired counts will occur with probability at most $1-\alpha$, where α is the desired confidence. By determining these unlikely splits, and referring to the observations, it is possible to set confidence limits for the population median. While strict confidence requires taking blocks of paired counts with at most a certain probability, interpolation can usually be carried out with safety.

Example 4 (See Figure 3). The differences between reaction-times of an individual and a control group were 6, 5, 3, 2, 1, −1, −2, −3, −5, −12, −13, −13, −15, −28, when expressed on a logarithmic scale (actually in terms of 100 times the log to the base 10). Within what limits do we have 90% confidence that the median of the population lies?

Drawing the 50–50 split, and parallel lines at ±8.4 mm., we find that the triangles for (10, 4) and (4, 10), where $4+10=10+4=14$ (the number of cases), are cut by these parallel lines about $\frac{3}{4}$ of the way from their vertices nearest the 50–50 split. We interpolate $\frac{3}{4}$ of the way from 1 to 2 (these are the fourth and fifth values from the top and $\frac{3}{4}$ of the way from −12 to −13 (these are the fourth and fifth values from the

bottom) to obtain approximate 90% confidence limits of $-12\frac{3}{4}$ and $1\frac{3}{4}$ for the median of the population of differences from which the 14 observed differences were a random sample.

Critique of Example 4. The interpolation is based on the fact that the use of 1.001 (say) for a cutting point would give the paired count (4, 10), as would any cut up to 1.999, while 2.001 would give (3, 10). If we take account of the grouping and rounding process we should widen these interpolated values by $\frac{1}{2}$ the grouping interval. Thus if the difference were rounded from more decimals to the nearest tenth, and were actually 6.0, 5.0, 3.0, $\cdots$, -28.0 then we should use $-12\frac{3}{4}-\frac{1}{20}=-12.8$ and $1\frac{3}{4}-\frac{1}{20}=1.8$. If, as might more reasonably have been the case, they had been rounded to integers, we should use $-12\frac{3}{4}-\frac{1}{2}=-13.25$ and $1\frac{3}{4}+\frac{1}{2}=2.25$.

COMPARISON AND ANALYSIS OF VARIANCE

Comparison

The comparison of two variances calculated for samples from normal populations, leads alternatively to Fisher's z or Snedecor's F. As explained in Section 14, the distribution of these quantities is mathematically related to the binomial distribution. Thus we may use binomial probability paper to make a significance test of the equality of the variances of two normal populations—the only approximation being the approximation of binomial probability paper to the binomial distribution.

The test is made by drawing the *line* or *split* through the point whose coordinates are the observed *sums of squares* of deviations from means, and plotting the *point* whose coordinates are half the numbers of degrees of freedom.

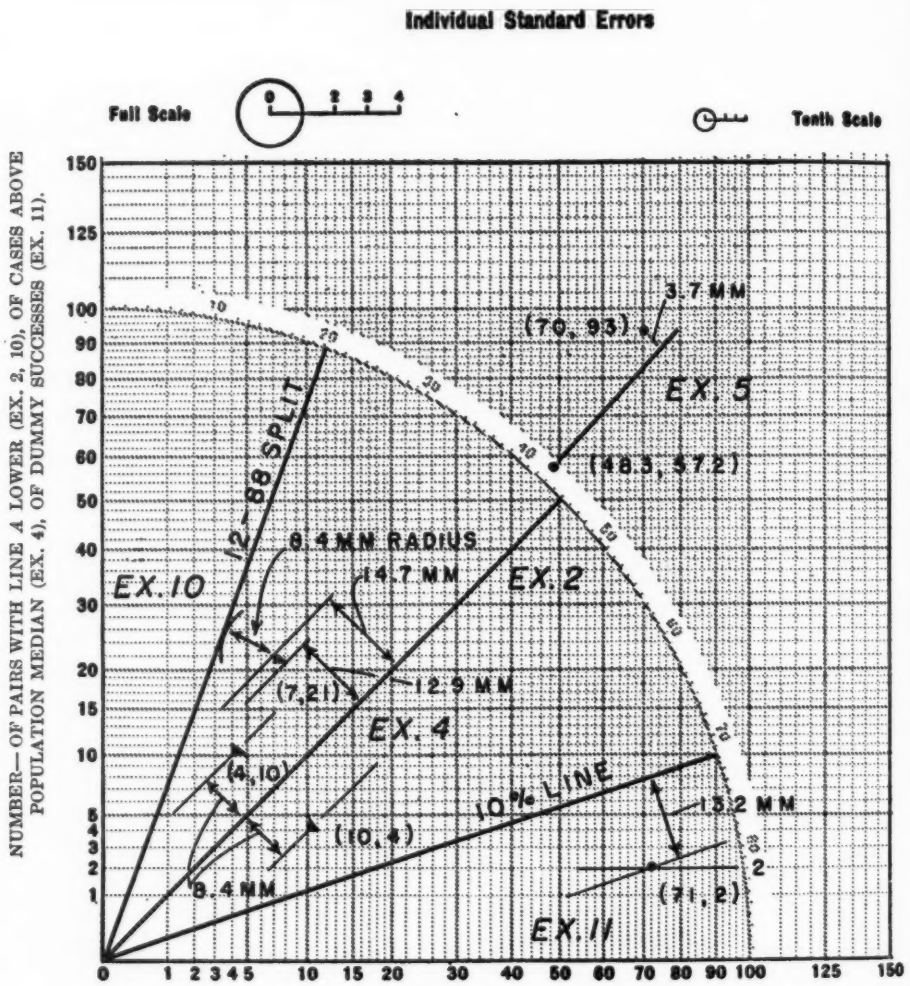
Example 5 (See Figure 3). In volume 1 of Biometrika, Fawcett [8, 1902, p. 442] gives the sample variances of the lengths of 141 male Egyptian skulls as 34.2740 and the variance of the lengths of 187 female skulls as 30.5756. Are these significantly different?

The sums of squares of deviations are 4832.6 and 5717.6. The degrees of freedom are 140 and 186. Drawing the split through (48.3, 57.2) and plotting the point (70 = 140/2, 93 = 186/2), we find a distance of 3.7 mm which is near a (two-sided) significance level of 45%. Thus there is no evidence of a difference in variance.

Critique of Example 5. The dangers in such a test lie far more in the possible lack of normality of the populations of practical experience than in the approximations of binomial probability paper. Such comparisons of variances from *independent* samples based on normality

FIGURE 3

SHOWING THE SIGN TEST (EX. 2), CONFIDENCE LIMITS FOR MEDIAN (EX. 4), THE *F* TEST (EX. 5), OPERATING CHARACTERISTIC OF SIGN TEST (EX. 10) SAMPLE SIZE FOR POPULATION TOLERANCE INTERVALS (EX. 11)



should almost always be taken with a grain of salt, particularly when so many degrees of freedom are involved.

This procedure is quite convenient when both degrees of freedom are greater than 24 (this is outside the range of short tables of F or z) and will be quite accurate if both degrees of freedom are at least 2. When one of the degrees of freedom is 1, accuracy will be improved by plotting $\frac{1}{2}$ seven-tenths of the way from zero to one, because $\sqrt{.5}=.7$. The use of binomial probability paper has the feature, interesting to some, of producing estimates of intermediate levels of significance. If it is desired to check the tabled values of levels of F against the results given by binomial probability paper, one plots the split (n_1F, n_2) and measures the distance to the point $(n_1/2, n_2/2)$, remembering that the tabled values provide a one-sided test.

Analysis

The same process can, of course, be applied to the analysis of variance, in principle a special case of the comparison of variances, but in practice a whole realm of its own. The sum of squares column determines the line and the *df* column the point.

Example 6 (See Figure 2). How can the analysis of variance in Example 7 below be tested graphically?

Draw the line through (257.15, 73.16) and plot the point (1, 5). The perpendicular distance of 15 mm is very highly significant (beyond the 0.1 per cent point).

Critique of Example 6. The difficulties with normality, mentioned in the last critique, are usually reduced enough to be neglected in an analysis of variance situation. Only in case this graphical method yields borderline significance is accuracy an excuse for using an F or z table, though convenience may often be a reason.

The angular transformation

The analysis of variance of counted data is frequently facilitated [3, Bartlett 1947 and references cited therein] by making the angular transformation. If the data involve small numbers, the accuracy of graphical transformation will suffice. The observation is plotted, a line through the origin produced to the quarter-circle, and the corresponding angle read off.

Example 7. W. E. Kappauf and W. N. Smith (personal communication) tested the performance of six observers in reading three types of dials. Sixty readings were made by each observer on each size. The

errors, angles, and analysis of variance are shown below. (The reader will find it easy to check on the computation of the angles on his own piece of graph paper.)

	Errors in 60 trials			Corresponding angles		
Observer	Size I	Size II	Size III	Size I	Size II	Size III
B_0	7	4	4	20.2	15.0	15.0
B	15	6	3	30.2	18.6	13.1
K	10	7	5	24.1	20.2	16.9
R	28	20	16	43.2	35.6	31.2
S	11	8	9	25.2	21.6	23.0
T	25	17	15	40.0	32.2	30.2

Source	df	Sum of squares	Mean square
Observers	5	991.64	198
Sizes	2	257.15	129
Interaction	10	73.16	7.3
Binomial	—	—	$\frac{821}{60}=13.7$*

* Variance of an angle obtained from binomially distributed data $\frac{1}{4n}(\text{radians})^2=\frac{821}{n}(\text{degrees})^2$.

The usual test for significance of the effect of size would be $F=129/7.3=17.8$ on 2 and 10 degrees of freedom, which is highly significant.

Critique of Example 7. A possibly more conservative test would be $F=129/13.7=9.4$ on 2 and ∞ degrees of freedom, which is still very highly significant. Although 7.3 on 10 *df* is not significantly less than 13.7, there is reason to believe in this case that the error mean square in a large-scale repetition of this experiment might well be less than 13.7. For the analysis above assumes the errors distributed binomially, and the probable differences in difficulty of the various dials attempted might reduce this variance.

PART III. APPLICATIONS TO DESIGN

BINOMIAL DESIGN

Since binomial probability paper allows us to approximately judge the significance of a paired count, it must also let us plan binomial experiments to have desired properties.

Sample size necessary to resolve two given percentages

When designing a test to discriminate between two theoretical percentages, the experimenter often wishes to know that *any* result will give significant evidence *against at least one of the two theories.* The procedure is best described by an example.

Example 8 (See Figure 2). A geneticist wishes to test whether a certain character appears in one-half or one-quarter of the progeny of a certain mating. He requires significance at the (two-sided) 1 per cent level against at least one of these hypotheses and wishes to know the smallest sample size which will guarantee this. He draws the 50-50 and 25-75 splits and then parallel lines at a distance of 13.1 mm (2.58σ) which corresponds to the two-sided 1% level. These parallel lines intersect at (37, 63). This point separates the triangle (36, 63), (37, 63), (36, 64) from the triangle (37, 62), (38, 62), (37, 63). Thus the paired count (36, 63) is beyond the 1% level from the 50-50 split, while (37, 62) is beyond the 1% level from 25-75. Thus a sample size of $36+63=37+62=99$ $(=37+63-1)$ will be enough.

Critique of Example 8. The design of this experiment is very good as far as sample size and significance levels are concerned. Since such genetical ratios are usually well-behaved, the experimenter who uses a sample of at least 99 and protects the progeny against causes of differential mortality should obtain very good results.

Our criticism should be directed against his less careful competitor who says that he will use the two-sided 1% level also, but will be satisfied with a sample size for which an observed 75-25 proportion will be at this level. Since the parallel line for 50-50 cuts the 25-75 split at (6, 18), he will use a sample size of only $5+18=6+17=23$. This design probably does not meet his needs, for if he uses it over and over he will be well protected from falsely stating the ratio is not 50-50 (since he is using a two-sided 1% level) *but he will miss one-half of the cases where the ratio is* 25-75. Such designs emphasizing one risk are usually ill-chosen, and if such a choice is compelled by limited experimental resources the choice of significance level should be reexamined, looking at both types of risk.

Designing single sampling plans

An essentially equivalent problem arising in industrial work as well as in certain kinds of experimental work is to design a sampling inspection plan which will distinguish between two kinds of quality, say product which is $100p_1$ per cent defective and that which is $100p_2$ per cent defective ($p_1 < p_2$). The plan desired is often described in terms of the operating characteristic curve, namely that large lots having $100p_1$ per cent defectives should be accepted $100(1-\alpha)$ per cent of the time, while lots having $100p_2$ per cent defectives should only be accepted 100β per cent of the time ($1-\alpha > \beta$). The process of building such a plan is described by the following:

Example 9 (See Figure 2). It is desired to construct a plan which will accept product with 3 per cent defectives, 95 per cent of the time, while product 12 per cent defective is to be accepted only 10 per cent of the time. What sample size should be used, and how many defectives can be tolerated before the lot is rejected?

We construct the 3% line (the 97-3 split) and the 12% line (the 88-12 split). If we wish to accept lots which are 3 per cent defective 95 per cent of the time, we must accept lots whose samples, as plotted on the paper, go as high as 1.64σ above the 3 per cent line, consequently we draw a parallel line $1.64\sigma = 8.4$ mm above the 3% line. Similarly in order to reject material which is 12 per cent defective 90 per cent of the time, we must reject lots whose samples come within 1.28σ of the 12% line, so we draw a parallel line $1.28\sigma = 6.5$ mm below the 12% line. The intersection of the two construction lines is (61, 5) and we find that the sample size is $61+5-1=65$, the acceptance number is 4, the rejection number is 5.

Critique of Example 9. It is assumed here that the lot is much larger than the sample, say ten times as large. Notice that these significance levels (10% and 95%) were one-sided. Consultation of tables shows that this plan will accept 3%-defective lots 96.77% of the time and will accept 12%-defective lots 9.69% of the time, which is approximately the result requested.

The operating characteristic of the sign test

We often want to know how well the sign test will discriminate.

Example 10 (See Figure 3). In Example 2 we considered 28 pairs of observations and decided to treat (7, 21) as significant. It is natural to inquire what population percentage of favorable pairs is needed to insure significance at this level 95% of the time. We must then find a

split so that the triangle representing (7, 21) is 8.4 mm $=1.65\sigma$ away. This leads to the 12-88 split, and so the sign test with 28 pairs discriminates very well between 50% and 12% (or 88%).

As in Examples 8 and 9, we can determine a sample size so that the sign test will have given discriminating power.

TOLERANCE LIMIT DESIGN

Sample sizes for population tolerance limits

In industrial work it may be desirable to take the least sample from an unknown population, such that the range from the smallest value in the sample to the largest value in the sample will cover a given fraction a of the population with given confidence β. This may be shown to be equivalent to a binomial problem, namely: Find the least sample size from a q-p split ($p=1$-q) such that the second count will be at least 2 with confidence β. (This is, of course, a special case of Example 8.)

If it is desired to use the rth from the bottom and the mth from the top to establish tolerance limits, replace 2 by $r+m$.

Example 11 (See Figure 3). A manufacturer of ball bearings wishes to have 99.5% confidence that 90% of his ball bearings lie between the limits set by the largest and smallest of a sample of a chosen size. He draws the 10% line (10% = 100% − 90%!), a parallel line $2.58\sigma = 13.2$ mm lower (for 99.5% confidence) and the horizontal line through 2. The intersection of the last two lines gives (71, 2) and the desired sample size is $74=72+2$.

Critique of Example 11. This example assumes that the successive ball bearings produced behave like a random sample—no manufacturer of any metal object, even ball bearings has reached so high a state of control. The practical interpretation of such a sample size is that it is a lower bound.

Second sample tolerance limits

When tolerance limits are desired for a second sample, rather than for the population, the problem is hypergeometric, and may only be approximated by a binomial problem. The chance that the range between the rth from the bottom and the mth from the top from a sample of n will omit N_1 or less from a second sample of N, may be shown to be the same as that a sample of N_1+n_1 (where $n_1=r+m$) from a finite population split N to n will contain n_1 or more of the second sort. If $N_1 \leqq N/10$, this can be approximated by the chance

of a second count of n_1 or more in a sample of N_1+n_1 from a N-n split.

Example 12 (See Figure 2). A manufacturer of precision resistors tests random samples of 1000 of each new type, and establishes the second from each end as working limits. He wishes to know the confidence with which he may expect (a) 99.5 or (b) 99 per cent of a batch of 50,000 similar resistors to fall within these working limits. He draws the 50,000-1,000 split and computes the chances of getting a second count of 4 or more in a sample of (a) 250+4, (b) 500+4. These are given by the distances from the split to (250, 5) and (500, 5), which are 0.0 mm$\approx$50.0% and 9.3 mm$\approx$3.4%. He has, therefore, 50.0% confidence that 99.5% will be within working limits and 96.6% confidence that 99% will be within working limits.

In the light of this example, the reader can construct answers to the other problems of tolerance limits for a second sample.

Critique of Example 12. If the manufacturer's production line is nearly in control when it starts to produce a new type, the authors will be surprised. The procedure given will answer the manufacturer who "wishes limits to the confidence, PROVIDED the process were in perfect control from the start." Also note that m and r have to be selected before the sample is examined, if the probabilities are to be accurate.

ANALYSIS OF VARIANCE DESIGN

Operating characteristic of model II—anova

An analysis of variance situation is Model II [6, Eisenhart 1947], when the effects are drawn from a normal population with variance σ_1^2, the errors being drawn from a normal population with variance σ^2. If the effects are the c column effects in a simple design with c rows and r columns, then the mean squares for columns and for error have the expectations $\sigma^2+r\sigma_1^2$ and σ^2 and the degrees of freedom (c-1) and (c-1) (r-1). In Model II, the mean squares are still distributed like multiples of chi-square. Thus the power of such an experiment can be easily determined as in the following example.

Example 13 (See Figure 5). A random sample of 25 sailors are to have their balancing ability measured quantitatively on each of 7 days. The results will be submitted to the analysis of variance, and a 5% significance level used. How large must σ_1^2/σ^2 be, before the existence of differences between sailors will be detected 95% of the time?

Here $c=25$, $r=7$, and the degrees of freedom are 24 and 144, hence we plot the point (12, 72). A one-sided 5% level corresponds to 8.4

mm, so we draw a circle around (12, 72) with this radius. The tangent splits are 27.2-100 and 8.9-100, so that the critical variance ratio is found from

$$\frac{\sigma^2 + 7\sigma^2}{\sigma^2} = \frac{27.2}{8.9}$$

to be $\sigma_1^2/\sigma^2 = .29$.

The basis of this construction is as follows: A ratio of sums of squares of 27.2 to 100 is needed for the chosen significance level. A ratio as small as 8.9 $(\sigma^2+7\sigma_1^2)$ to $100\sigma^2$ can be expected by chance 5% of the time if the population ratio is $\sigma^2+7\sigma_1^2$ to σ^2, hence to obtain 95% confidence of finding significance at 5%, we must have

$$\frac{8.9(\sigma^2 + 7\sigma_1^2)}{100\sigma^2} \geqq \frac{27.2}{100}.$$

PART IV—SEVERAL PAIRED COUNTS

LEVEL AND HOMOGENEITY COMBINED

Comparing several sets of data with a theoretical proportion

A set of several paired counts may give evidence against a fixed theoretical proportion in two ways. If the observed proportions are too variable, they indicate lack of homogeneity, that the samples came from populations with different proportions. If the average observed proportion deviates too much from the theoretical, it indicates a change (or error) in level, that the theoretical proportion does not apply.

It is well known that the correct way to test a homogeneous set of paired counts for agreement with a preassigned population proportion is to add them together, and then test the sum (as in Part II). Tests of homogeneity alone are the subject of the next section. Many delicate testing procedures involve first a test of homogeneity and then a test of level. Combined tests are mainly used to make quick and easy tests.

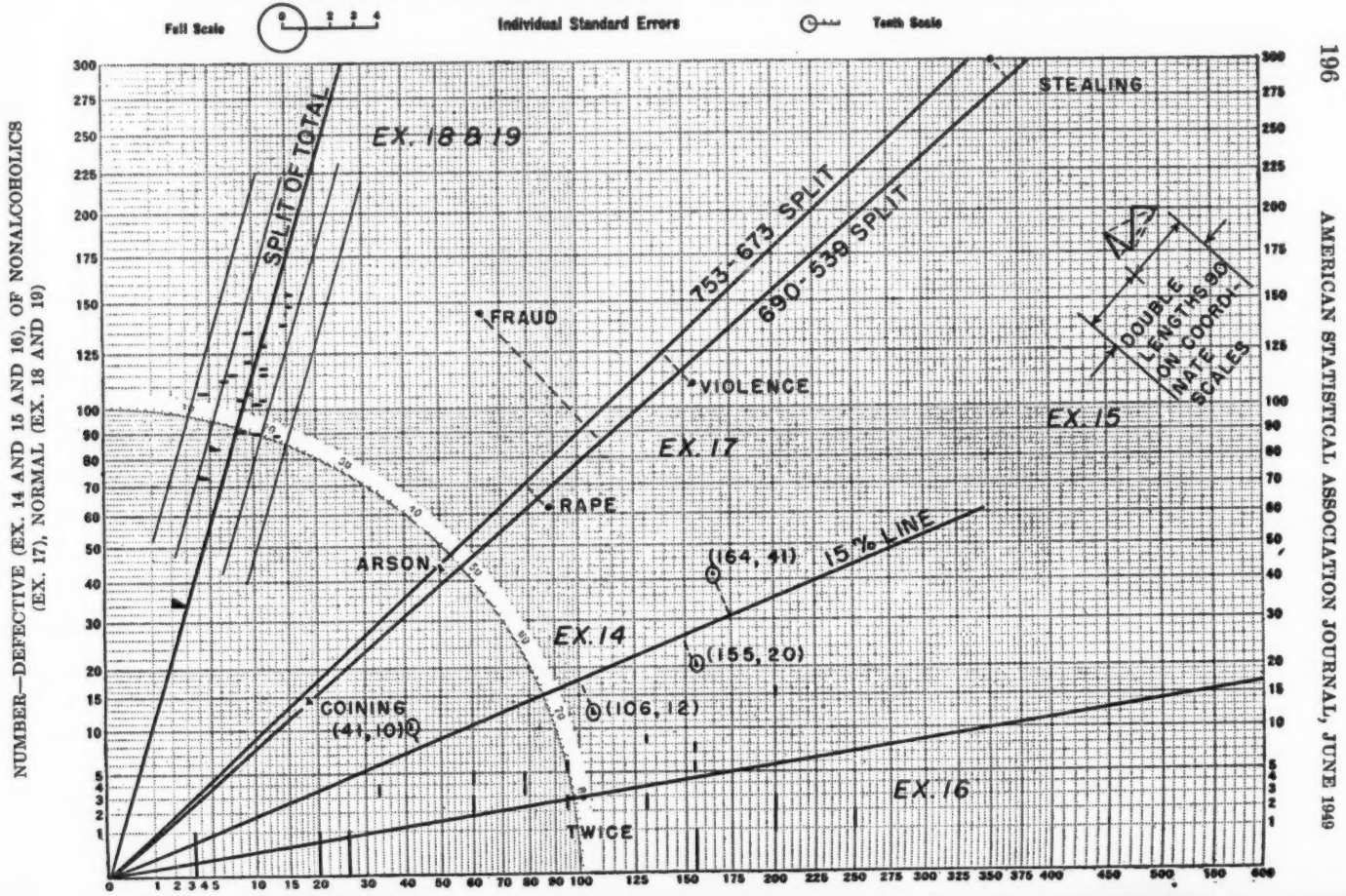
Example 14 (See Figure 4). A production process has been producing an average of 15% defective pieces over a long period. After the introduction of a new batch of raw material, successive shifts produced the following paired counts of nondefectives and defectives:

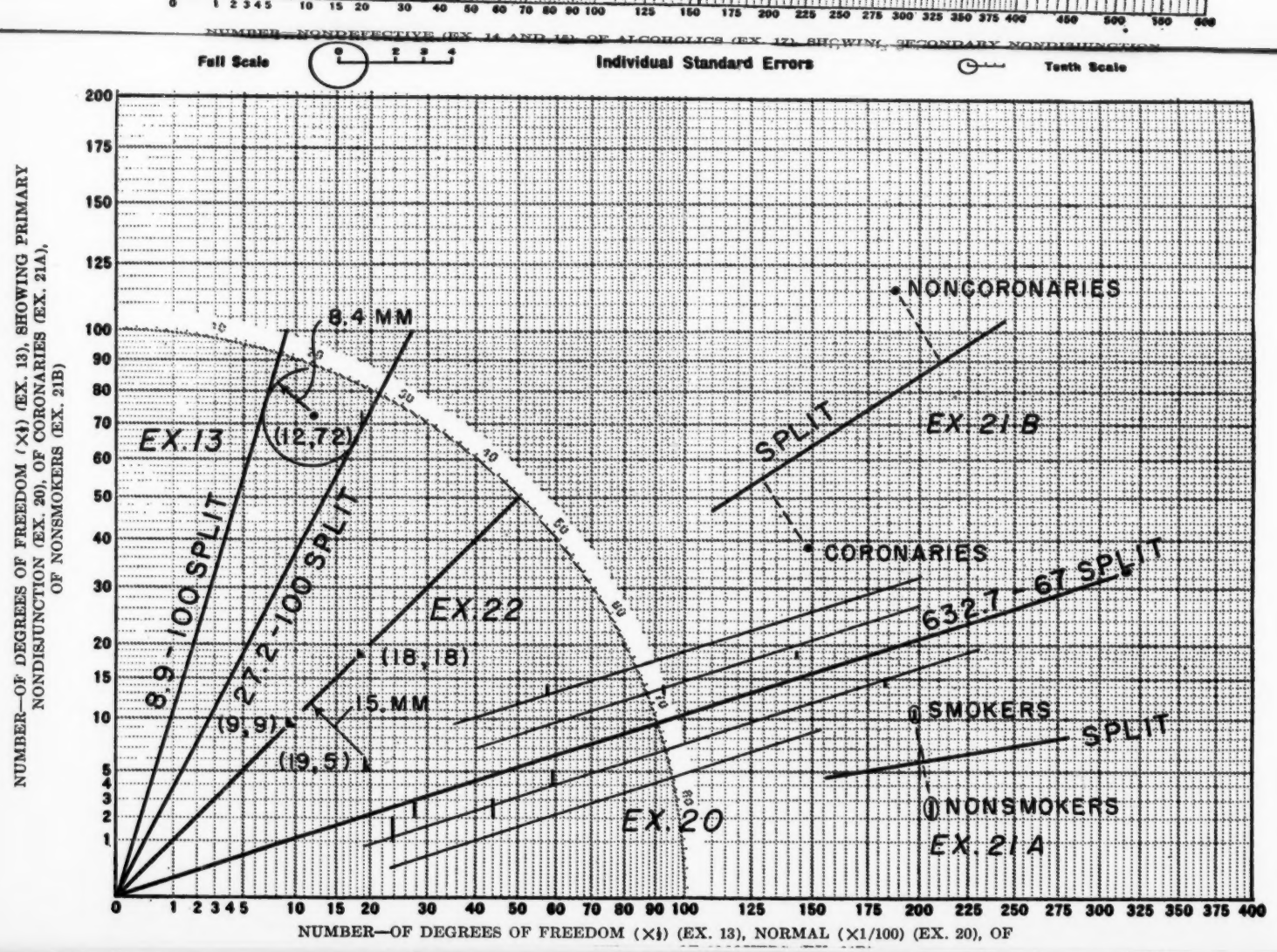
1. (155, 20), 2. (164, 41), 3. (106, 12), 4. (41, 10).

Is it reasonable to think that the production process is producing the same proportion of defectives as before? We plot the points and the

FIGURE 4

SHOWING COMPARISON OF K PROPORTIONS WITH THEORY (EX. 12, EX. 15), FIRST STEP TOWARD A STABILIZED p-CHART (EX. 16), HOMOGENEITY OF K PROPORTIONS (EX. 17, EX. 18, EX. 19).




Full Scale
Individual Standard Errors
Tenth Scale
NUMBER—OF DEGREES OF FREEDOM (×½) (EX. 13), SHOWING PRIMARY NONDISJUNCTION (EX. 20), OF CORONARIES (EX. 21A), OF NONSMOKERS (EX. 21B)
NUMBER—OF DEGREES OF FREEDOM (×½) (EX. 13), NORMAL (×1/100) (EX. 20), OF
8.4 MM
EX.13
(12,72)
8.9-100 SPLIT
27.2-100 SPLIT
EX.22
(18,18)
15.MM
(9,9)
(9,5)
EX.20
NONCORONARIES
SPLIT
EX.21B
CORONARIES
6327-67 SPLIT
SMOKERS
SPLIT
NONSMOKERS
EX.21A

15% line. Since no middle distance from the line is as great as $2\sigma = 10$ mm, and since the points are about equally spread about the line, there seems to be no reason to suppose the new batch of raw material has made a change in the percentage defective.

If some points were found outside 2σ the resulting paired count (of points inside 2σ vs. points outside 2σ) would be compared with the 19-1 split by the method of Example 2.

An alternative procedure is to combine the middle distances from the theoretical line by crab addition (on page 179) and consider this a chi-square with as many degrees of freedom as there are paired counts.

Example 15 (See Figure 4). Taking the same data as in Example 14, the crab sum, which can be obtained in less time than it takes to read the description, is a length, which when doubled may be read on the marginal scale as $\chi^2 = 9.0$ which is between the upper 10% and upper 5% points for 4 degrees of freedom.

Critique of Examples 14 and 15. Merely examining to see whether any points fall outside the 2σ limits does not squeeze the data dry, and is not very precise, but is very convenient in situations where a control chart would seem reasonable and proper. Both tests tend to detect either a change to a new level or excess variability due to changes in level from shift to shift. To check one of these alone, different procedures should be used. A change to a new level should be tested for by combining all the shifts, and thus comparing (466, 83) with 85-15 (which is far from significance). Changes in level from shift to shift should be tested for by the methods of the next examples (Ex. 17–19).

The stabilized p-chart

Where the lot size is constant in 100% inspected industrial production or the sample size is constant when sampling inspection is used, one of the standard quality control procedures is the p-chart, where the percentage of each lot or sample found to be defective is plotted against lot number or time.

When lot size is not constant, and it is not feasible to break down the data into groups of different size, there is a need for a new technique. The use of groups of varying size is not recommended—it would be better to break up the lots into rational sub-lots of nearly uniform size—but where the quality control engineer cannot arrange for the better solution he may wish to use the following device which was suggested to us by Acheson Duncan [5]. (The classical device is

to plot observed percentages, with broken horizontal lines for control limits. This makes a hard-to-read, messy diagram.)

Example 16 (See Figures 4 and 6). The following data on adjustment irregularities of electrical apparatus appear in the ASTM *Manual on Presentation of Data* [1, 1940 *Supplement B*, p. 58]. The first number given is the sample size, the second, the number of defectives. (We shall not correct the sample size by subtracting the number of defectives, because in the one case where this might make a visible difference there are no defectives.) We divide the sample size by 10, and plot the triangles in Figure 4 (to this scale, the triangles reduce to vertical segments).

ADJUSTMENT IRREGULARITIES, ELECTRICAL APPARATUS

Lot	Sample Size	Defectives	Lot	Sample Size	Defectives
1	600	2	11	1550	7
2	1300	2	12	950	2
3	2000	1	13	950	5
4	2500	1	14	950	2
5	1550	5	15	35	0
6	2000	2	16	330	3
7	1550	0	17	200	0
8	780	3	18	600	4
9	260	0	19	1300	8
10	2000	15	20	780	4

From Figure 4 we measure the vertical deviations from the $\bar{p}$ line (which is assumed to be 0.27% based on past experience) and plot them on a regular control chart (Figure 6), being sure to keep the data in the order in which they originally appeared. (The use of tracing paper makes this process very easy.) In practice each new observation would first be plotted on binomial probability paper (perhaps at an enlarged scale) and then transferred. If the data are retained on the original probability paper, the advantage of examining the data for trends and runs would be lost.

Critique of Example 16. The control chart in Figure 6 looks very different from that usually given. In the usual chart Lot 19 is shown as beyond the control limits on the high side, and Lots 4 and 7 are not detected as being possibly too defective-free because we find that there can be no lower control limit. This kind of plotting might be useful even when the samples are of constant size.

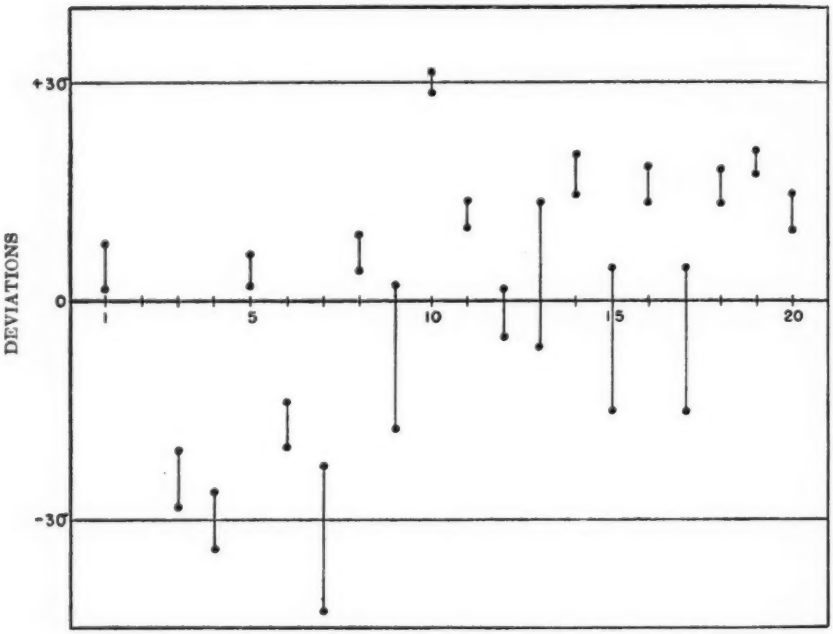
TESTS OF HOMOGENEITY

The general case

Given 5 or 50 or 500 paired counts which we wish to test for homogeneity, to test if it is reasonable that they have arisen from sampling

FIGURE 6

SHOWING COMPLETED, STABILIZED *P*-CHART AT ENLARGED SCALE (WITH SEGMENTS MADE VERTICAL).



a population with the same percentages in the two categories, the problem is the same, but the practical solution is different. In every case we plot the individual paired counts and draw the best fitting split, either by eye or through the sum of the paired counts.

We shall discuss three methods here, namely:

(1) graphical chi-square,
(2) range,
(3) counts in $\pm 1\sigma$ and $\pm 2\sigma$ strips.

Each of these has advantages and disadvantages, and, to the best of our knowledge, they can be compared as follows:

Method	Number of Samples		Advantages	Disadvantages
	Feasible	Recommended		
χ^2	2 to ∞	2 to 8 or 15	efficiency	labor
range	2 to ?	2 to 20	ease and speed	limited efficiency
counts	15 to ∞	15 or 20 to ∞	80% efficiency; relative simplicity	

The range is only recommended for 20 paired counts or less since its use for larger k involves the delicate details of the normal distribution and since its efficiency is less than the counting method.

To apply the χ^2-test, plot the paired counts and the split through their sum, and combine the *middle* distances by crab addition as explained at the end of Section 4.

Example 17 (See Figure 4). The following classic data by C. Goring quoted by K. Pearson and by M. G. Kendall compare the number of alcoholics and non-alcoholics among criminals according to crimes committed. The first number in each pair is the number of alcoholics:

Arson	(50, 43)	Stealing	(379, 300)
Rape	(88, 62)	Coining	(18, 14)
Violence	(155, 110)	Fraud	(63, 144)
		Totals	(753, 673)

The graphical display shows very clearly that (1) the observations are discrepant, (2) the crime of fraud is the only one for which the proportion of alcoholism is really different. Graphical chi-square computation gives 30.2 on 5 degrees of freedom—highly significant. When criminals convicted for fraud are removed from consideration the remaining five groups are each less than one standard deviation away from the new fitted line (690–539 split). Stealing is slightly misplotted.

Critique of Example 17. If the definition of "alcoholic" were sufficiently objective, and if the sample of convicted criminals represents a random sample of criminals, then the analysis seems sound. It cannot, of course, throw any appreciable light on the connection between alcoholism and crime in general, bearing only on the question "exclud-

ing fraud, do alcoholics tend to be convicted of some types of crime and non-alcoholics of others?"

A quicker method of analysis, and one well suited for drawing lines by eye is to compute the range of the sample, that is the sum of the greatest middle distances to the right and left of the line. Then the range measured in millimeters or standard deviations can be compared with Table 4 (at end of paper) to discover whether the samples deviate enough among themselves to provide evidence that the observations did not arise from random sampling from a single proportion in the population.

Example 18 (See Figure 4). In testing the effect of the X-chromosome inversion B^{M1} on secondary non-disjunction, K. W. Cooper (personal communication) raised 21 cultures of *Drosophila melanogaster*, crossing $v\text{-In}(1)B^{M1}/y^2w^av/Y$ females with wild-type Canton-S males. The presence or absence of secondary non-disjunction can be detected in female progeny. The 21 cultures gave the following results:

(9,135), (7,115), (11,118), (13,89), (15,148), (8,91), (6,113), (11,104), (9,122), (10,90), (15,155), (14,138), (5,84), (11,128), (2,34), (4,73), (4,107), (9,107), (10,103), (11,115), (8,104)

where the smaller class showed secondary non-disjunction. The split is drawn through the total of (192, 2273). The range is 16.5 mm which is far from significant—thus this test gives no evidence of heterogeneity.

Critique of Example 18. Clearly an amount of heterogeneity large enough to affect the estimated standard error of the grand mean would almost surely have been detected. The test seems adequate for its purpose.

A more refined, but still simple test is obtained by drawing parallel lines, at ± 5 mm and ± 10 mm. In sets of 21 or more samples, we expect about 5% outside the outer lines and about 33% outside the inner lines. The weighted sum

$$12\,(\text{no. outside 10 mm}) + 3\,(\text{no. between 5 mm and 10 mm}) - 2\,(\text{no. inside 5 mm})$$

is distributed with mean nearly zero and variance nearly $11.72k$ in case of k homogeneous samples. Approximate significance levels are $5.65\sqrt{k}$ and $8\sqrt{k}$ for the 5% and 1% points.

Example 19 (See Figure 4). Returning to the data of Example 18 and drawing the 5 mm and 10 mm lines, we find (classifying borderline cases according to the center of the hypotenuse of the triangles).

	Expected	Found
Outside	1	0
Between	6	4
Inside	14	17

The weighted sum is $12-34=-22$ which is far from the significance levels of 25.9 and 36.5. This more delicate test finds no evidence of heterogeneity in the per cent of secondary non-disjunction in Cooper's cultures. Indeed, if anything the data are a little too homogeneous, though not enough to notice. These methods can also be applied to the case of unsymmetric counts, as in the following example.

Example 20 (See Figure 5). In testing the effects of X-chromosome inversions on primary non-disjunction, K. W. Cooper (personal communication) crossed 847 males with females of eight different chromosomal sequences. Exceptional cases can be detected in both males and females. The observed counts for males were (2885, 13), (7172, 18), (4672,13), (9162,14), (1389, 2), (2961,4), (2199, 2), (1195, 1). Does the rate of primary non-disjunction seem to be constant?

The total count is (31635, 67), and the corresponding split together with the individual counts are plotted on Figure 5, where *all horizontal coordinates have been divided by 50.* The lines parallel to the total split are at ± 5 mm and ± 10 mm *vertically.* The range, measured *vertically* (since the horizontal coordinate has been reduced), is 19.3 mm which is not far from the 5% point of 21.8 mm.

Critique of Example 20. The method of Example 18b should not be applied on so few points, but if it were applied the weighted sum would be $12(1)+3(3)-2(4)=13$. The value of $5.65k$ is 16.0, but direct calculation of the 5% point yields 19. Thus the approximate 5% point is not too accurate for 8 points (16.0 is about the 10% point). Calculation of χ^2 by crab addition of the *vertical* deviations from the line to the total yields 13.9, which is again not quite at the 5% level for 7 degrees of freedom.

The four-fold table

The four-fold table, where a sample is classified into two categories in each of two ways has received very much attention by both applied and theoretical statisticians. Different methods of analysis have been given, some of which assume that

(1) the sample is a representative of samples in which only the total is fixed, or

(2) the sample is a representative of samples in which one set of marginal totals are fixed, or

(3) the sample is representative of samples in which all marginal totals are fixed.

Many of the "control group versus experimental group" experiments so common in biology, medicine, psychology, and education fall under (2), since the numbers in the control and experimental groups are fixed. Such experiments can be approximately analyzed as a homogeneity test as in the last section. For the case of *two* paired counts, the chi-square and range methods are equivalent, and the range is simpler.

Example 21 (See Figure 5). English et al. [7, 1940] took samples of 208 smokers and 208 non-smokers and investigated the incidence of coronary disease. They found (198,10) and (206, 2), where coronaries are the second category. The range is 17 mm which is significant at the 5% level.

They also took 187 cases with coronary disease, and 302 without, and investigated the incidence of smoking. They found (149,38) and (115,187), where smokers are the first category, which yields a range of 30.1 mm (Figure 5) which is horribly significant.

Critique of Example 21. These last two samples can be united into a four-fold table, but, in view of the way in which the data were obtained,

	Coronary disease	
	Yes	No
Smokers	149	187
Non-smokers	38	115

it would be *incorrect*, to compare (187,149) with (115,38) by this method and to assume that two binomials were being compared. However, the range obtained in this way is 29.7 mm and it is possible that such inverted tests on binomial probability paper give approximately correct answers.

A less obvious example

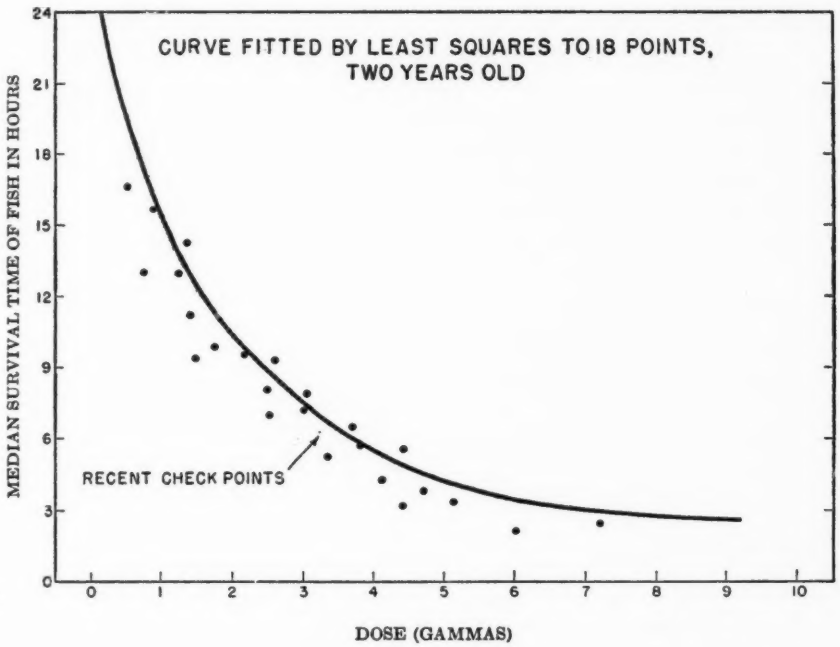
The ideas behind the sign test may be extended to give approximate tests in many situations of greater complexity. Such tests may be very useful, when used with the knowledge that they are quick, but often lack the sensitivity of more complex methods.

Example 22 (See Figures 7 and 5). A routine bioassay had been in

use for two years using a standard curve. Occasional checks on the standard had been made. The situation is shown in Figure 7, which raised two questions: (i) Does the curve agree with the recent points? (ii) If not, has something surely changed in two years, or may the difference be assigned to the combined sampling fluctuations in establishing and checking the standard curve?

FIGURE 7

BASIC BIOASSAY DATA FOR EXAMPLE 22 (A LESS OBVIOUS EXAMPLE).



The first question is answered by the split test, for comparing (19, 5) with a 50-50 split yields (Figure 5) a separation of 15 mm which is very highly significant.

The second question can be approximately answered as follows: the original least square fit to 18 points was probably more accurate than fitting a median to 18 points and less accurate than fitting a median to 36 points. A roughly fair test should come between a comparison of (19, 5) and (9, 9) and a comparison of (19, 5) with (18, 18). These give (Figure 5) ranges of 14.5 mm and 16.5 mm, which are both beyond the

5% level, indicating that the activity, or the fish, or the technique has probably changed.

Critique of Example 22. While these are not the most thorough tests which can be applied to this situation, anyone familiar with bioassay computation will appreciate their speed, simplicity, and clarity.

PART V—REFERENCE MATERIAL

MODIFICATION OF THE ANGULAR TRANSFORMATION

The original angular transformation

The angular transformation was introduced by R. A. Fisher in 1922 [9, p. 326] in a genetic situation where a certain proportion was varying by random fluctuation from generation to generation. In 1936, Bartlett [2, p. 74] proposed its use on experimental data as a means of stabilizing the variance when binomial data were subjected to the analysis of variance. Various authors have proposed its use for various purposes, a considerable number of references may be found in [13, 1947].

Bartlett's modification

In his 1936 paper, Bartlett also proposed an empirical modification to make the transformation more effective near $p=0$ and $p=1$. This was the device of transferring $\frac{1}{2}$ of a unit from the larger count to the smaller count. Thus (3, 29) would become (3.5, 28.5). This proved to be helpful, but had the annoying feature that both (3, 4) and (4, 3) were converted to (3.5, 3.5) which did not seem appropriate.

The smooth version

The smooth way of obtaining the good effects of $\pm\frac{1}{2}$ near the ends and ± 0 in the middle is to add $\frac{1}{2}$ to each cell, thus passing from $(n\text{-}k, k)$ to $(n\text{-}k+\frac{1}{2}, k+\frac{1}{2})$. It is clear that for values of p near 0, $\frac{1}{2}$, and 1 this will stabilize the variance very well. How well requires a numerical study, now in progress.

Correction for continuity

Most of the applications of binomial probability paper discussed above deal with tests of significance rather than with scoring paired counts. We must try, then, to assign nearly normal deviates, not to single paired counts, but to tails—to all $(n\text{-}k, k)$ for which $k \geqq r$, for example. This is closely connected with the scoring problem, since a natural dividing line between $(n\text{-}r+1, r\text{-}1)$ and $(n\text{-}r, r)$ is $(n\text{-}r+\frac{1}{2},$

r-$\frac{1}{2}$), and in accordance with the last paragraph, this is to be scored as if it were (n-r+1, r). Thus we expect to find that

$$\sqrt{n+1}\left(\sin^{-1}\sqrt{\frac{r}{n+1}} - \sin^{-1}\sqrt{p}\right)$$

is nearly the normal deviate associated with the probability that $k \geqq r$, where k is binomially distributed according to n and p.

Flattening

Since the angles involved are rather small, it is plausible to replace them by their sines. This is of course what has been done in the examples, where we have always measured distances perpendicular to the splits. A little trigonometry shows that the distance from (n-r+1, r) to the p-q split is (in standard deviations)

$$2(\sqrt{p(n-r+1)} - \sqrt{qr}).$$

(To obtain distances in millimeters, replace 2 by 10.16 mm.)

Accuracy

The accuracy of the over-all approximation to

$$Pr\ \{k \geqq r \mid k \text{ binomial } (n, p)\}$$

by

$$Pr\ \{x \geqq 2(\sqrt{p(n-r+1)} - \sqrt{qr} \mid x \text{ unit normal}\}$$

has been studied numerically, and a note giving details will be submitted to the Annals of Mathematical Statistics (by Murray F. Freeman and John W. Tukey). The general conclusion is that the approximation is extraordinarily good near the 1% to 5% points, and remarkably good in general.

THE INCOMPLETE BETA AND F DISTRIBUTIONS

The binomial distribution is, as is well known, given by the expansion of

$$(q+p)^n \qquad q = 1-p$$

where n is the number of cases, p the chance of a "success" and the term

$$\binom{n}{r} q^{n-r} p^r$$

in the expansion is the probability of exactly r successes. The probability of r or more successes is given by

$$S = \sum_{x=r}^{n} \frac{n!}{x!(n-x)!}(1-p)^{n-x}p^x.$$

Using the well known device of differentiating both sides with respect to p and summing, we get

$$\frac{dS}{dp} = \frac{n!}{(r-1)!(n-r)!}p^{r-1}(1-p)^{n-r}.$$

Replacing p by t and integrating S from 0 to S, and t from 0 to p, we have the usual relation

$$Pr\ (r \text{ or more successes}) = I_p(r, n-r+1)$$
$$= \frac{n!}{(r-1)!(n-r)!}\int_0^p t^{r-1}(1-t)^{n-r}dt$$

where $I_p(m, n)$ is the incomplete Beta-function. Hence if binomial probability paper successfully represents the binomial distribution it also successfully approximates the incomplete Beta-function.

Thus

$$I_p(r, n-r+1) \sim Pr\ \{x \geqq 2(\sqrt{p(n-r+1)} - \sqrt{qr}) \mid x \text{ unit normal}\}$$

which seems, incidentally, to be a new analytic approximation to the incomplete Beta-function. Simplifying notation, we find that $I_p(m_1, m_2)$ corresponds to the distance from p-q split to the point (m_1, m_2).

The ratio of two independent mean squares obtained from normal variates of the same variance is Snedecor's F, which is related to Fisher's z by $F = e^{2z}$. The ratio of the numerator sum of squares to the total sum of squares may be written in terms of F as

$$x = \frac{n_1F}{n_2 + n_1F},$$

and its distribution is given by,

$$Pr\ \{x < p\} = I_p(\tfrac{1}{2}n_1, \tfrac{1}{2}n_2).$$

Hence, to the approximation of binomial probability paper, a ratio

$$\frac{s_1}{s_1 + s_2} = r$$

of *sums of squares* has a probability of arising from populations of equal variance which is given by the tail area corresponding to the deviation of the count $(\frac{1}{2}n_1, \frac{1}{2}n_2)$ from the line $p-(1-p)$ which is the same as the line $s_1 - s_2$.

PART VI—INDEX, OUTLINE AND TABLES

Introduction

Table 1 is not intended to replace the worked examples, but rather to serve as a key for the new reader and a reminder for the old.

The short tables which follow are of standard distributions based on the normal distributions. Since millimeters are convenient units for use with binomial probability paper, they are given in both millimeters and in standard deviation units.

For maximum accuracy, use a sharp pencil! (Regular thickness automatic pencils may serve for some routine work, but finer lead will give better results.) The figures have been drawn for clarity, not accuracy.

Remember these methods are all approximations.

TABLE 1

INDEX AND OUTLINE

Example	Aim	Plotting Required	Remarks
		Part II. Plotting one observed quantity	
1 p. 182	Observed and theoretical proportions	1 paired count 1 split (theory)	Use short distance for significance level. Use both short and long for significance zone.
2 p. 184	Sign test	1 split (50-50) 1 paired count	
3 p. 186	Confidence limits for proportion	1 paired count 2 splits (at distance)	Use short distance
4 p. 186	Confidence limits for median	1 split (50-50) 2 paired counts (at distance)	Use short distance
5, 6 p. 187, 189	All F tests	1 split (sums of squares) 1 point ($\frac{1}{2}$ degrees of freedom)	
7 p. 189	Angular transformation	1 paired count 1 split (through middle)	

TABLE 1 (*Continued*)

Example	Aim	Plotting Required	Remarks
		Part III. Application to design	
8 p. 191	Designing binomial experiment	2 splits (theory) 2 parallel lines (at distance)	Distances correspond to one-sided significance levels at percentages to be controlled. (AQL and RQL=LTPD)
9 p. 192	Designing single sampling plan	2 splits (theory) 2 parallel lines (at distance)	
10 p. 192	Operating characteristic of sign test	1 paired count 1 split (at distance)	Use short distance
11 p. 193	Sample size for population tolerance limits	1 split 1 parallel line 1 horizontal line	Split-to-line distance desired confidence. Sum of counts-in determines horizontal
12 p. 194	Tolerance limits for second sample	1 split 2 paired counts (touching)	Split through 1st and 2nd sample sizes; distance to common vertex=confidence
13 p. 194	Operating characteristic of anova II	1 point ($\frac{1}{2}$ degrees of freedom) 2 splits (at distance) compute from ratio of split ratios	
		Part IV. Several paired counts	
14 p. 195	k proportions and a theoretical proportion	k paired counts 1 split (theory) 2 parallel lines (± 10 mm)	Expect 1 in 20 outside by middle distance.
15 p. 198	k proportions and a theoretical proportion	k paired counts 1 split (theory)	Combine middle distance by crab addition, (see p. 179)
16 p. 199	Stabilized p-chart	k paired counts split (assumed level)	Transfer to tracing paper as control chart
17 p. 201	Homogeneity of k proportions (k small)	k paired counts 1 split (sum)	Combine middle distances by crab addition (see p. 179)
18 p. 202	Homogeneity of k proportions (k 20)	k paired counts 1 split (sum)	Range of middle distances
19 p. 202	Homogeneity of k proportions (k large)	k paired counts 1 split (sum) 4 parallels (± 5 mm, ± 10 mm)	12 outside +3 between −2 inside 5%, 5.65 k; 1%, 8 k. Use middle distances.
20 p. 203	Homogeneity of k unsymmetrical proportions	as 17 or 19 with large count divided	As 17 or 19 with distances in undivided direction
21 p. 204	Four-fold table	2 paired counts 1 split (sum)	Range from middle distances

TABLE 2

MILLIMETER TABLE FOR NORMAL DEVIATE

Significance Level one-sided	Significance Level two-sided	Normal Deviate millimeters	Normal Deviate multiples of
50%	100%	0.0	.00
40%	80%	1.3	.25
30%	60%	2.7	.52
20%	40%	4.3	.84
16.5%	33.2%	5.0	.97
10%	20%	6.5	1.28
5%	10%	8.4	1.65
2.5%	5%	10.0	1.96
1%	2%	11.8	2.33
0.5%	1%	13.1	2.58
0.1%	0.2%	15.7	3.09
Conversion relation		5.080	1.000

TABLE 3

MILLIMETER TABLE FOR CHI-SQUARE

Degrees of Freedom	Undoubled Millimeters At an Upper Significance Level of			Multiples of σ^2 At an Upper Significance Level of		
	(50%)	5%	1%	(50%)	5%	1%
1	(3.4)	10.0	13.1	(.5)	3.8	6.6
2	(6.0)	12.4	15.4	(1.4)	6.0	9.2
3	(7.8)	14.2	17.1	(2.4)	7.8	11.3
4	(9.3)	15.6	18.5	(3.4)	9.5	13.3
5	(10.6)	16.9	19.7	(4.4)	11.1	15.1
6	(11.8)	18.0	20.8	(5.3)	12.6	16.8
10	(15.5)	21.8	24.4	(9.3)	18.3	21.2
15	(19.3)	25.4	28.1	(14.3)	25.0	30.6
30	(27.5)	33.6	36.2	(29.3)	43.8	50.9

TABLE 4

MILLIMETER TABLE FOR NORMAL RANGES

Number of Observations	Millimeters At Upper Significance Level of			Multiples of σ At Upper Significance Level of		
	(50%)	5%	1%	(50%)	5%	1%
2	(4.8)	14.1	18.5	(0.95)	2.77	3.64
3	(8.1)	16.9	20.8	(1.59)	3.34	4.10
4	(10.0)	18.5	22.2	(1.97)	2.65	4.38
5	(11.5)	19.6	23.2	(2.26)	3.87	4.59
6	(12.6)	20.5	24.1	(2.47)	4.04	4.74
7	(13.5)	21.2	24.8	(2.65)	4.18	4.87
8	(14.2)	21.8	25.3	(2.79)	4.29	4.98
9	(14.8)	22.3	25.8	(2.91)	4.39	5.07
10	(15.4)	22.7	26.2	(3.02)	4.48	5.15
15	(17.4)	24.2	27.6	(3.42)	4.79	5.44
20	(18.8)	25.4	28.6	(3.69)	5.01	5.64

REFERENCES

[1] American Society for the Testing of Materials, "Manual on Presentation of Data," Philadelphia (1940).

[2] M. S. Bartlett, "The square root transformation in analysis of variance," *Suppl. J. Roy. Stat. Soc.*, Vol. 3, 1936, pp. 68–78.

[3] M. S. Bartlett, "The use of transformations," *Biometrics*, Vol. 3 (1947), pp. 39–52.

[4] W. J. Dixon and A. M. Mood, "The statistical sign test," *Journal of the American Statistical Association*, Vol. 41 (1946) pp. 557–566.

[5] Acheson J. Duncan, "Detection of non-random variation when size of sample varies," *Industrial Quality Control*, Vol. 4 (1947–48) No. 4, pp. 9–12.

[6] Churchill Eisenhart, "The assumptions underlying the analysis of variance," *Biometrics*, Vol. 3 (1947) pp. 1–21.

[7] J. P. English, F. A. Willius, and J. Berkson, "Tobacco and coronary disease," *J. Am. Med. Assn.*, Vol. 115, pp. 1327–1328 (1940).

[8] C. O. Fawcett, et al., "A second study of the variation and correlation of the human skull, with special reference to the Naqada Crania," *Biometrika*, Vol. 1 (1902) pp. 408–467.

[9] R. A. Fisher, "On the dominance ratio," *Proc. Roy. Soc. Edinburgh*, Vol. 42 (1922) pp. 321–341.

[10] R. A. Fisher and K. Mather, "The inheritance of style length in *Lythrum salicaria*," *Annals of Eugenics*, Vol. 12 (1943) pp. 1–23.

[11] R. A. Fisher and K. Mather, "A linkage test with mice," *Annals of Eugenics*, Vol. 7 (1936) pp. 265–280.

[12] Frederick Mosteller and John W. Tukey (designers), *Binomial Probability Paper*, Codex Book Company, Norwood, Mass. 1946.

[13] Statistical Research Group, Columbia University, *Selected technique of statistical analysis*, McGraw-Hill, 1947.

TEACHING STATISTICAL QUALITY CONTROL FOR TOWN AND GOWN*

EDWIN G. OLDS
Carnegie Institute of Technology
AND
LLOYD A. KNOWLER
State University of Iowa

During the next five years, in manufacturing plants and in the engineering schools, there will be many new programs initiated to meet the demands for education in Quality Control by Statistical Methods. This paper has been written with the hope that it will be of some help in the planning and executing of such programs. Beginning with a description of types of courses and a discussion of possible content, it touches on the questions of who is to do the teaching and how the subject matter might be presented and motivated. The importance of follow-up work is stressed. The latter part of the paper discusses the choice and utilization of instructional aids and describes some of the materials now available.

I—TYPES OF COURSES

THESE REMARKS concerning the teaching of Statistical Quality Control will be confined to four general types of courses, each of which permits various subdivisions. The four general types will be referred to as: (1) intensive or so-called ten-day courses; (2) extension and evening courses which, by nature, usually are given on a part-time basis; (3) university or college credit courses; and (4) in-plant training courses. Naturally these types are not necessarily mutually exclusive.

Intensive Ten-Day Courses.

Although statistical work with special reference to applications in industry has been given in colleges and universities for many years, it is believed that it received a great impetus during World War II because of the thirty-odd so-called eight-day courses in Quality Control by Statistical Methods most of which were sponsored, to a large extent, by the War Production Board and the United States Office of Education under the Engineering, Science, and Management War Training Program in cooperation with various educational institutions. Repre-

* Presented at Section on Training of Statisticians of the American Statistical Association joint with American Society for Quality Control and Institute of Mathematical Statistics at Cleveland, Ohio, on December 27, 1948. Parts I and II were presented by Lloyd A. Knowler, Part III, under a separate title, was presented by Edwin G. Olds.

sentatives from government and industries engaged in production directly related to the war effort were permitted to attend these tuition-free courses. Many of the trainees in the short courses received, for the first time, any formal background in statistical procedures. Because of some continued need for the type of training which can be given in these courses, a few educational institutions have continued to sponsor similar courses on a moderate tuition or fee basis.

The basic course has generally been expanded to ten days. A typical ten-day intensive course has the first day set aside for executives. The Executives' Session is devoted to an explanation of the aims and possibilities of a quality control program and procedures for the installation of such a program in the plants is indicated. In addition to the executives the session is also attended by the trainees who expect to remain the entire ten days.

The program during the remaining nine days consists of a series of conferences, lectures, and laboratory periods. In addition to daily sessions, three or four evening sessions are scheduled. Instructors in the courses are continually bombarded with questions during the short intermissions between sessions. General lectures are given to the trainees. The discussion of specific problems and the working of laboratory exercises is facilitated by dividing the general group into small sections. While the emphasis is to apply the statistical method to design, specifications, production, and inspection, the evening sessions and informal discussions during the day are often devoted to additional phases of statistical work.

The course content in the intensive courses is ordinarily divided into two parts: (1) control charts; and (2) acceptance sampling.

The use of control charts of the usual types ($\overline{X}$, R; p; np; c; and $\bar{c}$) as a production tool is stressed; in fact, the comparison of a control chart to a highway, the descriptions of a control chart as a picture, a newsreel, or as an advertisement of the workers' product is quite common.

In studying a control chart it is usually necessary to introduce the concept of a frequency distribution. Following this, or in connection therewith, a control chart is constructed based upon shop data. This is followed by a laboratory exercise and discussion. Thereafter demonstrations of the effect of changes in average or range are given with the use of chips or beads as indicated in the second phase of this report. This general pattern of instruction is altered occasionally to fit the particular needs of the trainees in the course.

The facts that new uses are being discovered constantly and that

additional concerns are making use of statistical procedures as more individuals are being trained are emphasized. It is pointed out that a single chart on a single machine for a single operator in a large corporation is essentially a one man company. The value of a control chart for short runs and its value when applied to new lines of production is more than noted. Quality and economy are stressed. The diagnostic value and the predictive value of determining when and where to look for trouble are also stressed.

The introduction of acceptance sampling in the basic course pertains largely to the Army Ordnance tables, the Dodge-Romig tables, and an introduction to sequential analysis together with a demonstration and discussion of the defects of 10 per cent sampling.

Some attention is given to the construction of a sampling plan, largely to bring out the value of an operating characteristic curve and an average outgoing quality curve. The use of a certified control chart in the manufacture of goods to replace acceptance sampling, or as another type of acceptance sampling, has been advocated by the presidents of at least two large corporations in this country who use and who realize the value of control charts.

In addition to a regular staff in charge of the course, three or four periods are given over to representatives from different industries using statistical quality control who discuss the practical applications in their own control problems.

It has been found fruitful, in some instances, to maintain a rather aggressive follow-up program including at least two two-day clinics in which the trainees attending the course meet with representatives of the instructional staff to discuss the application of quality control to specific manufacturing problems, to check on the correct application of statistical procedures, and to provide an opportunity for exchanging information. These follow-up meetings are exceptionally well attended. In some instances additional statistical topics such as correlation, analysis of variance, design of experiments, chi-square, and other tests of significance are considered. For the greater part, these additional techniques are taken up in section meetings which tend to form as a result of the short courses.

It has also been found desirable to give so-called advanced courses of eight days' duration. The topics included in a typical course are: a review of control charts; significance of differences; analysis of variance; correlation—linear and multiple; further aspects of acceptance sampling; further aspects of sequential analysis; chi-square; and use of calculating machines. An instructional pattern, similar to that used for

the elementary courses, is followed and also follow-up programs are maintained.

The usual experience in these intensive courses is to find that, initially, representatives from industries want and demand only so-called practical materials. After taking the course and making some uses of their procedures, they learn the value of statistics and then decide that the difference between practice and theory is rather fuzzy, if a difference even exists. In fact, many take the view that what works in practice does so because it has a reason and that the reason is the theory. It is these "practical" persons who ordinarily request additional "theoretical" work. In fact, they often ask that what might be a semester's or year's work in mathematical statistics be presented during one forenoon, possibly in one hour; not only be presented but be set out in such a way that they can go back to their plant and use it in the particular problem at hand.

Extension and Evening Courses.

Many of the extension and evening courses given once or twice a week, occasionally once a month, follow the same general pattern as the intensive courses although to some extent many of the discussions are a little more formalized. Naturally the material on each topic tends to be more self-contained to the extent that it is usually unnecessary to attend the first lecture or two in order to take something away from the later ones.

It is encouraging to note that enrollment tends to be maintained in these evening courses even though a year may be spent in one series. It is also encouraging to note that the same individuals find it profitable to repeat a series so as to get new ideas from different lecturers or instructors.

There is a trend toward offering an advanced, as well as an elementary, series of lectures.

University and College Credit Courses.

The content of university credit courses may not be so clearly defined. Because of the broad aspects of mathematical statistics, each person who is instructing a course of this type has had different experiences; it would be natural for them to draw upon their experiences to a considerable extent. It is believed, however, that through discussions such as this an approach to standardization will be effected.

The catalog description of such a course might read: "The theory and applications of that part of mathematical statistics used in main-

taining control of the quality of a manufactured product or of a service; in the construction and use of acceptance tests, and the associated concepts of the operating characteristic curve and the average outgoing quality"; or references might be made to "Elementary statistical methods and their application to industrial problems; construction and interpretation of Shewhart control charts; Sealy's modified techniques; Dodge-Romig and Army Ordnance Tables for acceptance sampling; quality assurance for sampling by measurement; introduction to sequential analysis; methods of correlation; elementary analysis of variance."

A course in Statistical Quality Control should be given in that college or division where it is accessible to the student—not merely possible for him to register. It is relatively easy to effect this result in some schools—in others a special problem may be created. In so far as engineering colleges are concerned, a course in Statistical Quality Control might be given during the junior, senior or graduate year. In such instances the student will already have studied through the calculus, hence many more of the problems can be approached directly. The student, through his shop experiences, has a better idea as to the working of a machine, if not the working of industry.

For that reason, a three semester hour course following the general pattern of the recent avalanche of books on Statistical Quality Control, with such supplementary reports as the instructor desires, may orient the average student about as far as is necessary. The developments and illustrations of various principles would naturally be supplemented by demonstrations such as will be pointed out in the third part of this report.

Many students will observe the desirability of more work. Additional work in mathematical statistics may be taken immediately or after a year or two of experience. The additional topics which could be covered in a subsequent course or courses will be of considerable value for the quality control engineer of a large company. In fact, it is very likely that it will be of value to the quality control engineer in a smaller concern. It is not unusual for executive officers of large corporations to seek a quality control engineer with a background in shop and only a little work in statistics. After some understanding, however, the reverse is true. The demand for well trained statisticians is well known today. Much of this demand has been augmented, however, by initial applications under the direction of open-minded industrial executives. In fact, it might be observed that these persons have been most effective promoters and have also made possible many interesting prob-

lems for investigation. They have not been backward in advocating the need in their company for a person with training equivalent to a Ph.D. in mathematical statistics. These executives are particularly insistent, however, that the quality control engineer with such advanced training spend considerable time in shops and in assisting in the applications. They have realized that the main reason many executives and supervisors do not use mathematical statistics is that they have little or no knowledge of it.

In-Plant Training Program.

The in-plant training program is a very important phase of statistical quality control. Some of the larger companies have held a two-day training program for their top executives in which they attempt to give a rather broad overall picture of the subject. A program to achieve such an objective might be as follows: an introduction to statistical quality control; construction and interpretation of $\overline{X}$, R control charts; discussion of additional types of control charts; introduction to acceptance sampling; and suggestions on putting quality control to work. Executives are very much interested in such a program. They are particularly interested in the demonstrations which seem to have some relation to their plant operations.

Following this two-day program, the executives then select an individual from each plant in their company to take a ten-day course either sponsored by the company or by an educational institution. If within the company, the material considered is that pertinent to their own problems. The quality control engineers then return to their separate plants and make a pilot run as an application. They then train persons of supervisory and operator level in those particular aspects in which those persons happen to be most interested and for which they feel the greatest need. This may be the use of control charts for production or for design, or it may be some aspect of acceptance sampling. In the meantime, the quality control engineers join various statistical societies, start a library, and continue their study. It is not unusual for them to seek personnel from the educational institutions to assist them.

The procedure of having the in-plant training program divided into several periods wherein classroom discussions of a week or two are followed by a week or two of work in the plant, and the process repeated, has a great deal of promise.

II—PREPARATION OF TEACHERS

Since statistical quality control includes so many other modern statistical techniques in addition to control charts and acceptance sampling, it might not be out of order, in a paper such as this, to make a few remarks concerning the preparation of teachers. It is recognized that much depends upon the individual in charge of the course. In the first place, it is essential that the teacher have a knowledge of the subject and be interested in transmitting that knowledge or in getting another individual to gain similar knowledge. As with so many professions, the work of a quality control engineer is getting beyond the cook book style. It is rapidly becoming essential that the prospective teacher study two or three years of mathematical statistics, with one or two of these years in the Graduate College. In order for him to do this with much facility it is desirable, if not necessary, that he have a rather thorough background in pure mathematics. Also, the prospective teacher needs a background in some area such as engineering, business or industry from which to draw information. This is a large order, to be sure, but we need to realize that teaching is a profession and that it should be so recognized, both in the preparation required and in monetary rewards.

III—INSTRUCTIONAL AIDS

In the previous sections we have discussed what material might be taught, how it might be organized and motivated, and who should do the teaching. In this section attention will be directed toward the working tools needed and how they can be used to clarify quality control principles and develop power in analyzing manufacturing problems. Teachers in many fields seem to be capable of presenting their subjects acceptably when equipped with nothing more than a basic text, a portable blackboard, a stick of chalk and an eraser. For many, in fact, this seems to be the maximum list of requirements. Such does not seem to be the case for statistical quality control where successful teachers find effective use for a large collection of various kinds of paraphernalia.

Classroom and laboratory equipment can be classified under five general headings: (1) Textbooks; (2) Material for supplementary reading; (3) Problems; (4) Calculating machines; and (5) Gadgets. Some brief comments on each may prove helpful to embryo teachers.

Texts.

To remark that the success of a course depends, to a considerable extent, on the choice of text is distressingly trite, but in this case, it

seems necessary. When the Office of Production Research and Development initiated its statistical Quality Control Program early in 1943, pitifully few books on the subject were available, and most of them would be classified as treatises, rather than texts. Since the war, publishers have been able to put on the market a considerable number of books with the Quality Control label, but it seems fair to assume that most of the books were not prepared for use as basic texts in the field. Before any of them are adopted for the classroom the following questions should be considered:

1. Does the book give a sound exposition of the general philosophy and statistical principles basic to statistical quality control?
2. Is the exposition of sufficient breadth and completeness to meet the needs of the course?
3. Is the level of sophistication appropriate, in view of the students for whom the course is planned?
4. Will the book generate enthusiasm for statistical quality control as an engineering tool?
5. Is the arrangement and problem content such that effective day-by-day assignments can be made?

It would be surprising to find many books meeting all five tests satisfactorily. A book written to sell Quality Control to the busy executive would miss its aim if it attempted more than a brief overview of the simplest methods together with a briefer explanation of why they work. A manual prepared for use in the shop seldom will contain any derivation of the formulas being applied or hint as to many of their limitations.On the other hand, a book directed at advanced undergraduate or graduate engineering students with calculus, and perhaps some elementary statistics "under their belts" would be tragic for an in-plant training course at the supervisory level.

The rapidity with which manufacturing organizations have adopted statistical methods for the improvement of their operations can be attributed, to a large extent, to the missionary efforts of industrial men who were trained in short wartime courses. Reports of savings in manpower and materials, together with improvement in quality spurred them to study the methods, to apply them to their own problems, and to spread the gospel. There is no diminution of need to whip up enthusiasm but, with the frenzied haste of war work at an end, there is more time for careful appraisal of success stories. Furthermore, the serious student is more likely to be impressed by the powerful nature of statistical methods than by dollars saved. In no sense does this

comment imply that case studies should be deleted from a text or course, but rather that they should be chosen cautiously and imbedded in enough detail so that the student has some opportunity to check the correctness of the solution.

Students differ in the amount of verbiage they need, but almost without exception, they need problems to clarify and emphasize principles. Problems daily, like apples, seem to have a most beneficial effect. Texts without problems, or with problems involving excessive amounts of mechanical manipulation and calculation, place a heavy burden on student and teacher. It does not follow that, simply because a set of data originated in a plant, it will be an efficient hammer to drive home a principle.

One further remark on the choice of a text. A successful book or course does more than arouse interest, inculcate principles, and develop skills; it must create and foster the urge for continuation of the learning process; and it must chart the path and locate some of the important land marks. In other words the book should contain a generous and appropriate list of references which will carry the student beyond its boundaries in all directions.

Material for Supplementary Reading.

Even if a text with a satisfactory set of references is located, it is too much to expect either that all useful material will be included or that all of the papers listed will be readily available. For those readers who have examined university, city or company libraries this lack of availability need not be elaborated. Therefore, the teacher of statistical quality control will find it necessary to compile his own list for outside reading and find a way of giving his students easy access to any unpublished reports which they should see. The authors of this paper are much relieved that there is no room in this paper to include recommendations regarding books to purchase. Fortunately, many of the leading statistical, mathematical and engineering journals not only list new books received, but also print authoritative and prompt reviews of many of them. At least two bibliographies on statistical quality control [1] [2] have been compiled rather recently. Both books and periodicals are listed, but the field of coverage is largely restricted to control charting and acceptance sampling. Furthermore, both will be somewhat out of date by the time this paper reaches publication.

The bi-monthly journal, *Industrial Quality Control*, is the official publication of the American Society for Quality Control and, naturally, caters to the needs of its members. These needs are not greatly differ-

ent from those of beginning students in the field and, therefore, the publication is a rich source of material. The *Supplement to the Journal of the Royal Statistical Society* also specializes in industrial applications of statistics, but at a somewhat higher level. Also, useful papers will be found in the *Journal of the American Statistical Association* and *Biometrika*.

The *Annals of Mathematical Statistics* is devoted, primarily, to basic theory, but some of the theory is directly applicable to quality control problems, even under the narrowest meaning of the appellation.

A set of twelve *Quality Control Reports*, prepared at the close of the war under the auspices of the Quality Control Program at Carnegie Institute of Technology, is still available and can be obtained at a cost of two dollars from the Office of Technical Services, Department of Commerce, Washington. In the main, these are case histories of initiation of quality control methods.

Excellent papers on statistical methods and their applications appear from time to time in many scientific, engineering and trade journals. These may pass unnoticed unless students are briefed to watch for and report them. Several of the societies hold occasional symposia on statistical methods. Usually the reports of these meetings are worth locating. Unpublished plant reports are useful, particularly if prepared by former students.

Sources of Problem Material.

There is a dearth of satisfactory problem material for easy reference. Most teachers of statistical quality control do not hesitate to beg, borrow, or even purloin all of the industrial data they can. But most files contain pitifully few really good problems. Perhaps "good problem" should be better described. Mr. Wyatt Lewis, of the Ontario, California plant of General Electric Company, contributed such a problem for inclusion in the *Outline Manual for Quality Control by Statistical Methods*, which was written by Working and Olds and used by them and by many others in the teaching of intensive eight-day courses during the war. (The *Manual* has been available for two dollars from the same source as the *Quality Control Reports* mentioned above.) The problem is called the Rheostat Knob problem and is an application of the Shewhart Control Chart techniques. Data on an easily described quality characteristic is given for several days' production for which both specifications and manufacturing conditions were changed. Interesting details are supplied. The problem divides naturally into several parts, each of which can be assigned as a separate

exercise. When a calculating machine is used or when the numbers are coded, the time for computation is not excessive.

The above-mentioned manual has a few other good problems. Manuals prepared for intensive courses given more recently at various universities have others. Recent books have some. Others can be manufactured by judicious use of the data and circumstances given in technical articles. Good problems *can* be dug up, but a modicum of work is required—and, perhaps, a little harmless chicanery.

Calculating Machines.

Recently one of the authors read about an elementary course in statistics which was said to have been given quite successfully ("not ideally!") without the use of a computing laboratory. A similar remark might be made regarding beginning courses in statistical quality control. Obviously it is debatable and would not apply to courses at an advanced level. Statistical quality control *can* be done without calculating machines, and there seem to be two principal arguments for such a procedure: (1) economy; and (2) the distressing fact that some industrial organizations do not furnish machines for their quality control departments and so their men need to learn how to get along without them.

For the most effective teaching of statistical quality control the authors agree in favoring a computing laboratory with sufficient equipment to accommodate a class of students. Not all classes would be held in the laboratory but, when used for a class, each member would have an automatic calculator of the same model. Needless to say, the laboratory should be kept open at stated hours for individual work, with an assistant in charge, competent to teach the operation of the machines and responsible for their ordinary care.

It is useful to have late models of the principal kinds of calculators available, so as to broaden the students' experience. Punch-card equipment is a welcome addition. Several of the gadgets to be described in the next section are a proper part of laboratory equipment. So are statistical tables and some reference books. If space and supervision is adequate, there is little danger of collecting too much equipment.

Gadgets.

In days of fairly recent yore, the average teacher of elementary probability reached the heights when he had each of his students toss a coin 100 times, recording the succession of heads and tails. Then he was content to retreat to a consideration of either the few reports of

such experiments in the literature or "cooked-up" examples of the same type. It is hard to understand why he failed to appreciate the pedagogical value of designing an experiment to illustrate a point of theory, predicting the result, running the experiment, and then taking the consequences if it turned out wrong. Whatever the reason, it is fortunate for the field of statistical quality control, that its leading teachers have broken away from tradition and have shown no hesita-

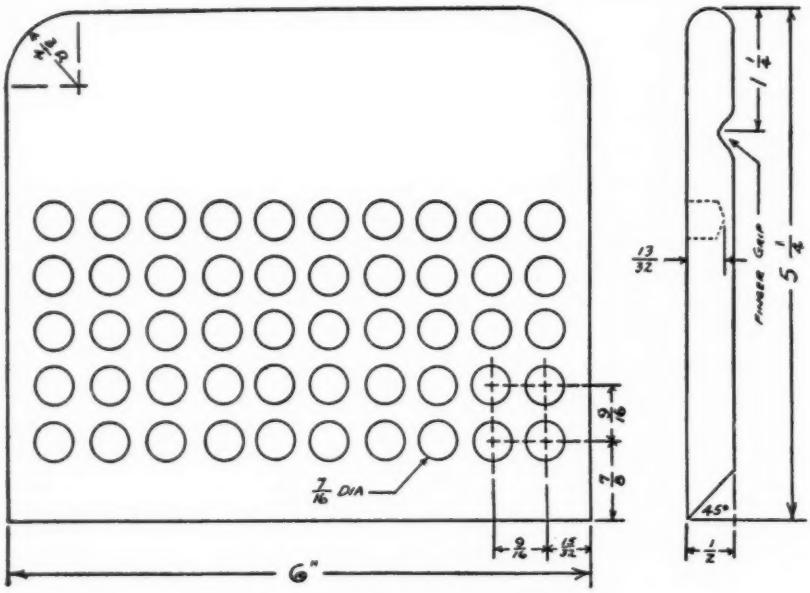


EXHIBIT 1
PLAN FOR 50 HOLE SAMPLING PADDLE
(adapted from a drawing by H. B. Rogers)

tion in using any available gadget which promised to assist in fixing a concept in the student's mind.

At some time during the last five or six years not only thousands of students, but scores of corporation presidents and their associates have been introduced to the vagaries of acceptance sampling by means of a box of beads and a 50-hole sampling paddle. The authors do not know who originated the paddle, but the man who made it famous was Holbrook Working. His paddle (see Exhibit 1) started as a wooden board about $6'' \times 6'' \times \frac{1}{2}''$. In it were sunk five rows of ten holes each, about 9/16″ apart, center to center, 7/16″ in diameter and 13/32″

deep. Near one edge the board was grooved to provide a handle. The rest of the equipment consisted of a cardboard box of several hundred 10-mm wooden beads (about 1200 white and 200 red). When Dr. Working went to the Walco Bead Company, 37 W. 37th St., New York City to purchase the beads and explained his purpose he met with astonishment, if not incredulity.

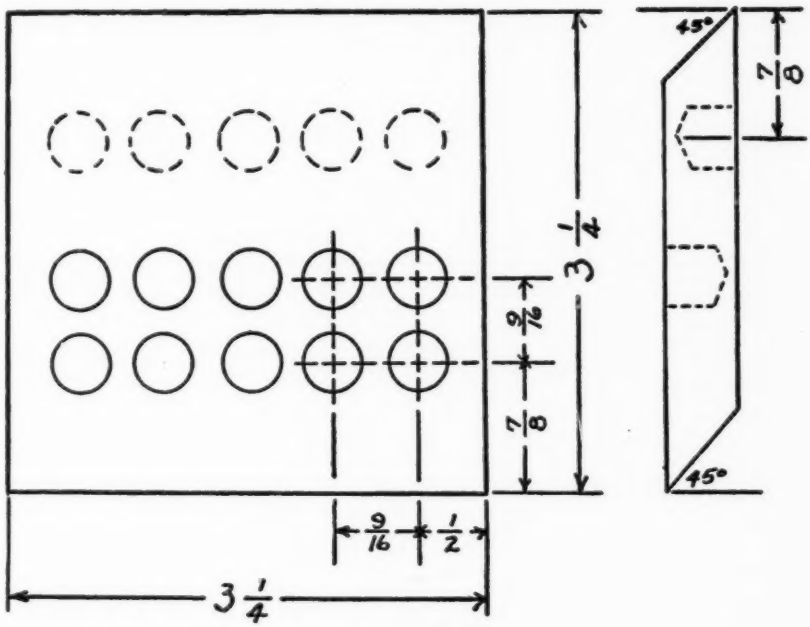


EXHIBIT 2
PLAN FOR 5 AND 10 HOLE SAMPLING PADDLE
(adapted from a drawing by H. B. Rogers)

The best known demonstration is concerned with illustrating the weaknesses of 10 per cent sampling. In the box 1152 white beads (good items) are mixed with 48 red beads (defective items). A paddle of beads (representing a lot from a controlled process) is withdrawn, the number of red beads counted, and the fifty beads dumped into a second box. Then, with a 5-hole paddle (see Exhibit 2), five beads are scooped from the fifty and number of defectives noted. The agreement is to accept the lot only if there are no defectives in the sample. Of course some of the lots accepted are worse than many of the lots rejected. Also the per cent defective in the uninspected portions of rejected lots

EXHIBIT 3

RECOMMENDED FREQUENCY DISTRIBUTIONS FOR SETS OF DISKS FOR SAMPLING FOR VARIABLES*

Set No.	1A	1B	1C	2	3A	3B	4
Color of Figures	Green	Black	Black	Blue	Black	Red	Red
Mean	0	0	0	+2	0	+1	+4
Standard Deviation	1.715	1.715	1.715	1.715	3.470	3.470	1.715
Numbers	Frequency						
−10					1		
− 9					1	1	
− 8					1	1	
− 7					3	1	
− 6					5	3	
− 5	1	1	1		8	5	
− 4	3	3	3		12	8	
− 3	10	10	10	1	16	12	
− 2	23	23	23	3	20	16	
− 1	39	39	39	10	22	20	1
0	48	48	48	23	23	22	3
1	39	39	39	39	22	23	10
2	23	23	23	48	20	22	23
3	10	10	10	39	16	20	39
4	3	3	3	23	12	16	48
5	1	1	1	10	8	12	39
6				3	5	8	23
7				1	3	5	10
8					1	3	3
9					1	1	1
10					1	1	
11						1	
	200	200	200	200	201	201	200

* It is necessary to have sets designated as 1A, 2, 3A and 3B, or sets differing little from these as regards means and standard deviations. Sets 1A, 2, and 4 will be thrown together for one demonstration, and bear figures in different colors to permit subsequent sorting. Sets 1B and 1C are used in a demonstration that calls for drawing from 5 bowls, 3 of which are alike.

The chips are white plastic discs, 9/16″ in diameter and 3/32″ thick. They can be purchased from Lamb Seal and Stencil Company, 824—13th Street, N.W., Washington, D. C. The wooden beads used in the sampling demonstrations are 10 mm. in diameter. They can be purchased from the Walco Bead Company, 37 W. 37th Street, New York City. A supply of 1200 white, 200 red, and 100 blue beads is suggested.

is about the same as the per cent defective in the uninspected portions of accepted lots, etc., etc.

It requires considerable restraint on the part of the authors to prevent them from devoting an entire paper to a description of the various uses of the paddle in connection with standard sampling tables, control

charts, tests of significance, confidence intervals and the like. A second paper could be written on "chips drawn from a bowl." White plastic chips about $\frac{5}{8}''$ in diameter and $3/16''$ thick are marked with numbers to simulate a normal distribution. Exhibit 3 gives suggested distributions for the various bowls which were found useful in O.P.R.D. courses. (These distributions seem to have been devised by Holbrook Working.)

Drawing samples of five from one of the bowls provides the data for a control chart. Changing bowls gives a graphic picture of the effect of a changed mean or standard deviation. With two bowls the meaning of Fisher's t-test becomes more clear. Correlate pairs of units and the presence of sample correlation when the universe correlation is zero is demonstrated in a fashion more telling than any appeal to theory. Use three bowls or more and analysis of variance can be introduced successfully. Possible demonstrations for stratified or sequential sampling are easy to plan and execute. The list of other useful demonstrations is almost unlimited.

Is it true that in random assembly, the square of the natural tolerance of an addition of several components is equal to the sum of the squares of the natural tolerances of the components? Professor MacCrehan has a set of 100 blocks to demonstrate that this is good theory. There are five sets of painted blocks; red, black, white, yellow, and green. In each set the twenty blocks vary in thickness. Assemblies are made against an upright board, on which are painted pairs of lines bounding the natural and absolute tolerances for the assembly. When five blocks, taken at random, one from each set, are piled against the board, the top of the pile falls between the two natural tolerance limits. Yet five blocks can be found which will reach the upper absolute tolerance line or just match the lower absolute tolerance.

A less spectacular, but none-the-less, useful gadget is a table of random sampling numbers. Several such tables are available [3], [4], [5]. While the same jobs can be done with a table as with the ships or beads, the use of a random sampling table at the elementary level does not seem to be very effective. At a higher level such a table is almost invaluable in connection with distribution theory. Suppose, for example, one one is forced to investigate the behavior of statistics in random samples from a very odd universe, one for which the density function might even be unknown. Numbers from the table can be so assigned that the universe is simulated in a form convenient for sampling. Then a large number of samples can be drawn easily and quickly, and the statistic under scrutiny can be calculated, tabulated, and studied. No further

work may be needed to arrive at satisfactory answers to many practical problems.

Having emphasized the point that gadgets are useful and lacking space for detailed description, the authors close this section with a list of several other physical aids to instruction. These gadgets, as well as those listed above can be used by students as well as teachers. In fact, students can learn some statistics by devising their own gadgets.

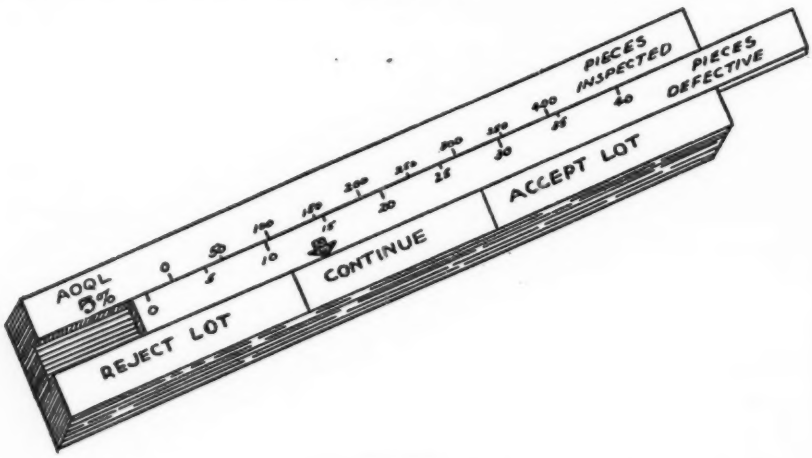


EXHIBIT 4

SKETCH OF A SLIDE RULE FOR SEQUENTIAL SAMPLING USED BY SERVEL, INC.
(from a photograph supplied by W. E. Gibbons)

1. Galton Board or Quincunx
2. Slot machines, seized in police raids
3. Sets of dice—fair and biased
4. Slides and movies
5. Colored plastic balls or ball bearings or marbles (These may be used with a paddle instead of beads. Both bearings and marbles are heavy and noisy.)
6. A can, for mixing chips, mounted on a phonograph turn table. The can is tilted so that the audience can observe the mixing procedure.
7. Pieces of wooden doweling of an assortment of lengths and diameters, with weights of pieces marked on them. This gadget is useful in explaining multiple correlation.
8. A sampling machine designed by T. H. Brown, and D. H. Leavens. Beads are mixed by revolving a closed container shaped like an

oil can. A sample is obtained by tilting the can so that beads roll into its transparent spout.

In addition to the gadgets built to provide for striking demonstrations of principles, the various mechanical devices concocted to help with practical application are deserving of mention. Exhibit 4 is a sketch of a slide rule being used effectively for sequential sampling. A slide-disk calculator has been designed for calculating standard deviations. Another type can be used in the shop to get control limits for average, range, and fraction defective charts. At least one quality control engineer inserted baffles in sheet metal cans of various sizes in such a fashion that they would mix lots of small parts and separate out random samples of specified sizes.

The authors have made no attempt to give an exhaustive list of gadgets. No doubt many readers have others of comparable merit. If so, it is to be hoped that careful descriptions of such gadgets, together with reasons for their construction and examples of their use, will be written and submitted for publication.

ADDENDUM

At the suggestion of a referee, who was present at the meeting when the above paper was presented and discussed, some brief comments are being added. They may delineate more clearly the nature of statistical quality control and indicate some of the main objectives of elementary courses in the field.

Statistical quality control is the application of statistical methods to the improvement of the manufacturing operation. At each stage in the life cycle of a product, from recognition of consumer need, through design, specification, fabrication and inspection, to final assurance of consumer satisfaction, statistical methods can play an effective role. All brands of quality control are concerned with the same problems but only *statistical* quality control utilizes *statistical* methods in solving them.

Industrial statistics and statistical quality control have many concepts and techniques in common but the dual classification is necessary. Many executives in charge of manufacturing regard industrial statistics as closely akin to business statistics and, therefore, mainly preoccupied with questions arising in sales and accounting. While these executives recognize the importance of such questions, they are primarily interested in problems of manufacturing. Therefore, they welcome any potential aid to quality control engineering which statistical methods may have to offer.

In a one-semester or two-semester course, or in intensive courses, there is neither the expectation nor the implication of transforming engineers into statisticians. The average engineer does not have the prerequisite mathematical background for much theoretical statistics. More preparation would be useful but it is a choice between taking him as he is or missing the opportunity to provide him with a few of the fundamental concepts and methods which he can grasp readily and apply with confidence. Most teachers try to pack as much statistics as possible into the time allotted. While statistical quality control should certainly reach beyond control charts and acceptance sampling by attributes, these two topics crowd a one-semester course. Parenthetically, this may be the reason why some engineers view statistical quality control as comprising only control charts and acceptance sampling.

In conclusion, it might be noted that there is little danger that a good teacher will encourage his students to view themselves as statistical experts. It seems to be more general to find students of statistical quality control very conscious of their limitations. This has two good effects: they are willing to seek advice and they are eager to learn more statistics.

REFERENCES

[1] *"Bibliography on Statistical Quality Control,"* Report No. 11 of a series of Quality Control Reports published under the OPRD Quality Control Program, Carnegie Institute of Technology, Pittsburgh, 1945.

[2] G. I. Butterbaugh, *"A Bibliography of Statistical Quality Control,"* University of Washington Press, Seattle, 1946.

[3] L. H. C. Tippett, *"Random Sampling Numbers,"* Tracts for Computers, No. 15, Cambridge University Press, London, 1927.

[4] M. G. Kendall and B. Babington Smith, *"Tables of Random Sampling Numbers,"* Tracts for Computers, No. 24, Cambridge University Press, London, 1939.

[5] R. A. Fisher and F. Yates, *Statistical Tables for Biological, Agricultural, and Medical Research*, Oliver and Boyd, London, 1938. Table 33 is a table of random numbers.

THE USE OF SAMPLING IN GREAT BRITAIN*

C. A. MOSER

Assistant Lecturer in Statistics, London School of Economics

The object of this article is to survey the use of sampling methods in Great Britain. The most important sample surveys undertaken by government departments, research organisations and commercial agencies are described with particular reference to their aims and sampling methods. Deficiencies in British sampling practice are discussed and suggestions for possible future developments are made.

INTRODUCTION

ONE OF THE most notable developments in the field of statistics in recent years has been the increasing utilisation of sampling methods in the study of human populations. Although a number of countries have shared in this development, American advances in techniques and applications of sampling have been the most striking—the more to be valued because they have been so fully described in numerous papers in the technical journals, more especially the *J.A.S.A.*

In Great Britain, the situation is rather different. Not only is the use of sampling more limited than in the United States but, for various reasons, the surveys that are carried out are rarely published. The result is that statisticians elsewhere have little opportunity of becoming familiar with the fields in which sampling is used in this country, and the methods employed.

The reports issued in conjunction with the meetings of the U.N.O. Sub-Commission on Statistical Sampling [1] indicate both the considerable use that is being made of sampling in many countries and the need for widespread collection of data and exchange of ideas on sampling methods. It is the purpose of this article to describe and comment on the more important sampling investigations carried out in this country by Government departments, commercial organisations and other bodies; and to discuss the main features of English sampling practice, pointing out deficiencies and possible developments in the use of sampling here. Discussion will be confined to investigations concerned with human populations leaving out of account, for instance, research carried out on agricultural experimentation, where sampling methods of a far more involved and refined character are employed. Throughout

* My thanks are due to statisticians in Government departments and elsewhere for making available to me much of the information on which this article is based.

the article, emphasis is on the fields of application and sampling procedures, rather than the results of surveys.

THE DEVELOPMENT OF SAMPLING IN GREAT BRITAIN

The present state of sampling in this country is best seen against the background of a long drawn out development. The first proper use of sampling techniques was in Professor Bowley's survey of Reading in 1912 [2]. He took approximately every 20th working-class household, paying great attention to the calculation of sampling errors and to the possibility of bias being introduced through substitution and refusals. This pioneering use of sampling by Bowley proved a great stimulus to social surveys, which had hitherto been based on non-random selection (Booth, 1889) or complete enumeration (Rowntree, 1901). All the major surveys of the inter-war years, such as the New London Survey, Merseyside, Southampton and others, follow to a greater or lesser extent the sampling methods first used by Bowley.

The first use of sampling in connection with official information was in John Hilton's enquiry into "The Personal Circumstances and Industrial History of 10,000 Claimants to Unemployment Benefit." Extreme care was taken to achieve a representative sample (the 10,000 workers were approximately a one per cent sample of all the claimants) and to avoid the many possible sources of bias. The sampling procedure was fully described by Professor Hilton [3] and, as F. F. Stephan points out in a recent paper [4], it is odd that his methods were not imitated by other Government departments. It was not, in fact, until 1937 that sampling was used in any large-scale Government investigation. In that year, the Ministry of Labour undertook its enquiry into working class expenditure [5] with the object of furnishing information on which a revised Cost-of-Living Index could be based. Budgets of expenditure for four weeks (spaced at quarterly intervals) were obtained from about 10,000 households—the initial selection of households having been made by sampling at random from various registers. The planned revision of the Cost-of-Living Index was interrupted by the war, but the new Interim Index of Retail Prices, which has taken the place of the old Index, is based on the results of the 1937/8 survey [6].

The 'thirties also saw the beginning of Listener Research and other Opinion Research bodies, but it was the war which, as in America, gave the decisive impetus to the utilization of sampling techniques. The need for quick and cheap information led to the foundation of an official Social Survey Unit and to other projects which will be described in this article.

THE NATIONAL LISTS

Perhaps the most important single difference between British and American sampling practice lies in the existence, in this country, of several lists covering the population; certainly, this is the key to what may appear to be simply lack of enterprise—our failure, so far, to make more use of modern sampling developments, and particularly area sampling. It will be seen that nearly all the major surveys to be described are based on samples selected from one or other of the lists. For that reason, and because the coverage and accuracy of the lists differ, a short note on each of them will be useful at this point.

(1) *The Maintenance Register*

This national register started in September 1939, when National Registration came into force. It is kept at local Food Offices and covers the whole population, with the exception of the Armed Forces and Seamen. The Register is a live one, in so far as cards of persons, who have moved out of or into a district are refiled fairly rapidly in the new district. There are separate files for persons under 16 and 16 and over, the cards being filed in order of code number (depending on the district where the card was initially issued).

Apart from the code-numbers, the card gives details of the person's name, sex, address and age (last birthday and date of birth). It must be said that not all the cards have been filled in with all these details and there is a certain amount of inaccuracy; but efforts are being made to bring the Register up to a higher standard of accuracy and completeness.

It will be seen that it is simple to draw a random sample of under 16's or 16's and over, or any other large age and sex group from this register.

(2) *The Ministry of Food Files*

The civilian Ration Books in this country last for a period of one year and there is an exchange of new for old books every July. At this time, the reference pages are extracted from the old books and are filed in alphabetical order of surnames at the local Food Offices. Unlike the Maintenance Register, the Ministry of Food file is only brought up to date once a year so that at any time during the year, the Register in any particular area will include persons who have died or have moved out of that area since last July. Dummy pages are inserted for persons moving into the area. The Register is divided into three separate groups:—under 5, 5–18, over 18.

(3) *The Rating List*[1]

The Rating ledger, kept by the Rating Officers of local authorities, is a list of all the rateable units in the area. It is generally in order of wards

[1] "Rates" correspond to U. S. local taxes.

and in alphabetical street order within wards; within streets, the order is simply by street-number. The entry for each dwelling unit (and, for rating purposes, flats are separate units) gives a description of the property (house, flat, cinema, shop, etc.) and the names of both the owner and occupier. As these lists are used for rating purposes, they are fairly accurate and up-to-date. By eliminating those types of property not required, it is clearly possible to form a sample of dwelling units. In order to sample successfully for households from this list, allowance must be made for the existence of multiple households and the difficulties arising from the distinction between households and dwelling-units.

(4) *The Electoral Rolls*

This is a list, published annually, of all persons entitled to vote in elections; that is, broadly speaking, British subjects aged 21 and over. It is arranged in the same order as the Rating List but, of course, gives not only the occupier of, say, a flat, but all persons in the flat who are eligible to vote. The pre-war list of 1939 was especially useful, in that the head of the household and his wife were indicated by separate symbols. On the new register, this distinction is not made. This register is not very accurate and, if employed at all for sampling purposes, it should be used in conjunction with and checked against the Rating List.

BRITISH SAMPLE SURVEYS

A discussion of sampling surveys falls conveniently into three parts. In the first place, a fairly full account will be given of official sampling investigations; then, some of the more important surveys undertaken by semi-public bodies and research organizations will be mentioned. Finally, an indication will be given of the work of Market Research and Public Opinion bodies.

A

OFFICIAL SAMPLE SURVEYS

(1) *The Social Survey*

The most important organisation in the Sampling Survey field is the Government Social Survey Unit. Founded in 1941, to undertake surveys urgently needed in the administration of Britain's war economy, it is now (as part of the Central Office of Information) a well-established Government research organization receiving an annual Treasury[2] grant of £60,000. The last two or three years have seen not only a considerable expansion in its size and its output—the Social Survey now employs some 250 field workers distributed throughout the country and a research and administrative staff of about 80—but also an im-

[2] The functions of H. M. Treasury are considerably wider than those of the U. S. Treasury Dept.

provement in the quality of its work, including its sampling techniques.

The position of the Survey unit in Government administration is best explained by indicating the procedure governing its work. A survey is planned in response to a request from a Government department, but is put into the field only after Treasury approval has been given. When the survey is finished, a report stating its methods and results is written by the Research Officer in charge, and is submitted to the department concerned for interpretation. While this procedure has the advantage that the Survey undertakes surveys covering a wide variety of subjects, it appears to involve certain drawbacks which are worthy of mention. These are essentially drawbacks arising out of the position, rather than the work, of the Survey.

a. The work of the Survey is largely confined to *ad hoc* investigations, the results of which are urgently required by Government departments. Consequently, not enough time is left for surveys which may be of interest more from a sociological than from a direct administrative point of view. More especially, as it is difficult to experiment with methods in a survey the results of which have been requested and are to be used by a department, there is not as much methodological research as appears desirable. It is to be hoped that the Social Survey will be able to spend more and more of its time on research into sampling techniques, interviewing methods, questionnaire biasses and all the other problems associated with social surveys.

b. The position regarding the publication of the results of the Social Survey is highly unsatisfactory. The reports are, of course, sent to the relevant department and many of them are made available on request at the offices of the Survey. The Survey of Sickness results are published by the Registrar-General and, every now and then, the results of a survey find their way into the Press. It is essential that the reports be given a wider circulation. They might, for instance, be offered for sale at H.M.S.O.

c. Thirdly, it is difficult to escape the impression that the conducting of the survey and the interpreting of its results are too widely separated. The Social Survey ought to be more than a mere collecting agency. The Government department usually keeps well in touch during the planning and execution of a survey but there should be more consultation between the department and the Research Officer in charge of the particular survey over the interpretation of the results and any actions arising directly from them.

It is, of course, impossible in this article to refer to many of the 150 or so sample surveys which the Social Survey has undertaken but it may be useful to name some of the more interesting investigations conducted for different departments and then to say something about the sampling techniques employed. The following list gives an idea of the scope of the Survey's work:—

For the *Board of Trade*

> Numerous surveys on wartime shortages of consumer goods; on effects of and attitudes to clothes rationing; on the use of demobilisation coupons; on public knowledge of the need for an active export policy, etc.

For the *Ministry of Food*

> Numerous surveys on different aspects of rationing; on the attitude to National wheatmeal bread and to National Milk and Cocoa, etc.

For the *Ministry of Information*

> Surveys on the public attitude to various films, books, posters and other publicity media.

For the *Ministry of Labour.*

> Surveys connected with the recruitment of women for industry; of miners and of agricultural workers.

For the *Ministry of Health.*

> The Survey of Sickness; surveys on the public reaction to the publicity campaign on V.D. and diphtheria immunisation; and on the public attitude to the nursing profession.

For the *Ministry of War Transport.*

> Surveys on workers' transport difficulties and on road safety.

In addition there have been regional surveys in Middlesbrough and Willesden; a survey on the impact of air raids; a large survey designed to give a picture of the distribution of the population and of the family and household composition throughout the country, surveys on the demand for holidays, shopping hours, the incidence of deafness, the employment of old persons and many more.

It is evident from this selection that the surveys undertaken by the Social Survey not only cover a very wide field, but also that they concern a large variety of populations, so that sampling methods vary a great deal. According to Box and Thomas [7]

> "The types of population sampled in Wartime Social Survey inquiries may be classified roughly into three groups. In many inquiries, information is required about the whole adult civilian population of Great Britain. A rather larger group of problems concerns only housewives . . . . A third type of problem relates to particular groups."[3]

A fairly typical Social Survey sampling scheme is that used in its

[3] It would now be more accurate to say that most of the surveys concern the adult civilian population.

Survey of Sickness. This regular monthly survey, which began in 1945, is an attempt to derive information about the incidence of all kinds of illness in the adult population. Each monthly sample of about 3000 adults (over 16) is obtained as follows:—The correct proportion of interviews is allocated (according to the current population estimate of the Registrar-General) to each of the twelve Civil Defence Regions. Within each Civil Defence Region, the administrative districts are divided into three groups:—

(a) Towns large enough (over 300,000) to be entitled to 30 or more interviews. (The Social Survey tries to arrange its samples so that about 30 interviews are allocated to each interviewer. This is found to be necessary from the point of view of cost and interviewer efficiency.)
(b) Other Towns.
(c) Rural Districts.

In Group (a), no further division is made. The Registers of all the towns are sampled (such towns account for about 10 per cent of the national sample). In Group (b), the towns are divided into groups by population size in such a manner that each size group is entitled to about 30 (or multiples of 30) interviews. The rural entitlement of interviews for the region is divided into units of approximately 30 interviews each. In both group (b) and (c), the requisite number of towns and of rural districts is chosen, as far as possible, at random, consideration being given to ensure that each of a number of geographical sub-regions receives approximately the correct quota of interviews. In all cases, the individuals for interview are taken at regular intervals from the Maintenance Register in the particular towns and rural areas selected. The over-all sampling fraction is about 1 in 10,000.

In other surveys, such as the investigation of water-heating appliances in domestic dwellings or that of crockery stocks, interest is in households rather than individuals. In these surveys, the samples were selected from the Electoral Rolls and the Rating List respectively. In yet other surveys, such as those on Road Safety, the Demand for Holidays and that on Shopping Hours, quota sampling was used.

It will be seen that so much variation in sampling practice exists within the Social Survey that it is difficult to give a complete picture. The following general points may, however, be made:—

a. In the surveys concerned with the adult population of Great Britain, the samples are usually between 1500 and 3000 cases.
b. For most of the surveys, one or other of the lists mentioned above form the basis of the sampling. The Maintenance Register and the Rating

List are the most frequently used, for the sampling of individuals and households respectively. In surveys connected with special populations, recourse is made to other records. So, for instance, in the Survey of Pneumokoniosis, the sample of 900 was taken from Ministry of Fuel records of men certified by the Silicosis Boards; in the survey on Blood Transfusion, the sample of 1200 donors was taken from local records of the Blood Transfusion Service.

c. The weak link in sampling schemes such as the one used in the Survey of Sickness lies in the selection of the actual towns and rural districts; as indicated above, this selection is largely governed by the desire to achieve complete geographical coverage of the area. But it is limited by considerations of cost and interviewer economy and efficiency. The samples are small, in terms of absolute numbers, and the attempt to allocate the interviews in bunches of at least 30 necessitates principles of selection which are not altogether satisfactory. It is believed that the Social Survey is giving thought to this problem.

d. It is satisfactory to note that the Social Survey is moving further and further away from the use of quota sampling which, at one time, was its chief method of sampling. It may be hoped that, ultimately, the method may be abandoned altogether.

The discussion of the work of the Social Survey has, inevitably, been short and incomplete. There is no doubt that the Survey has done very useful work and also, as its officers themselves would probably agree, that there is plenty of room for improvement in its methods. The time is ripe for a complete and up-to-date statement of the Survey's work on the lines of the paper presented in 1944 by Box and Thomas and, perhaps even more urgently, for a full and critical examination of all the techniques employed in its surveys.

(2) *The Ministry of Food*

The Ministry of Food was one of the first Government Departments to make any considerable use of sampling surveys. In addition to surveys undertaken for the Ministry by commercial organisations, the Social Survey has, during the last few years, carried out a large number of ad hoc investigations on different aspects of the country's food situation. The most important of the Ministry's various survey projects, however, is its continuous Family Food Survey, the aims and methods of which will now be outlined.

Object. The Family Food Survey is the chief source of information concerning the dietary habits of the population. Started in 1941 its objects are (a) to investigate "the nutritional value of food actually consumed as compared with the estimated physiological needs of the same families"; and (b) to collect data on the uptake of welfare foods

and on the effect of food control measures, including price changes. A household or family is defined as "all persons for whom the housewife caters".

Population. From the beginning of the survey the main sample has concentrated on households representative of the working class population of the country. In addition, similar middle class and rural working class samples have been taken more or less continuously for comparative purposes in most years. Special groups, such as miners, agricultural workers and old age pensioners are investigated from time to time.

The Sample. The Ministry take a fresh sample of approximately 800 households every month, so that about 10,000 households are covered annually. The sampling is done in two stages.

(1) Towns, including all the great conurbations, are first chosen by purposive selection—i.e. with regard to their size and the character of the region.

(2) Within the selected towns, sampling is done at random from the Electoral Register. In practice, the procedure is to tackle one part of the town, or ward, at a time so that interviewing is done as economically as possible. Predominantly upper class wards are excluded from the sample, as are those consisting mainly of industrial premises. Any middle-class households sampled in mixed wards are added into a separate middle-class sample. When lists for a whole ward are exhausted (the sampling fraction is 1 in 35), a new ward is started on. This continues until the whole town is covered; then a new and, if possible, similar town is chosen for sampling.

The Ministry of Food points out that the representativeness of the resultant sample is subject to two qualifications:

a. In the first place, small towns tend to be slightly under-represented, mainly as a result of the inevitable immobility of the interviewers.

b. In the second place, in the past about 20 per cent of the households sampled proved to be non-contacts even after three calls, while a further 30 per cent could not or would not co-operate for different reasons. When all else fails, interviewers are permitted to take substitutes by calling at each house in turn to the right of the one originally sampled.

This large-scale substitution is a potential source of bias and this must be borne in mind in interpreting the results. The fact that the sample has had a somewhat overweighted proportion of children may be attributable to this substitution. As far as average food expenditure per head is concerned, analysis has shown little difference between pre-selected and substituted households.

The fieldwork of the survey is undertaken by an outside commercial agency, so that direct contact between the Ministry and informant is avoided. Fieldworkers call every second day during the week of investigation or, if necessary, every day.

Sampling Variation. Another point to consider is the variability of the population in terms of family size, sex and age composition and the number of meals taken at home and outside. For this reason, as the Ministry points out, repeat samples taken at the same time and with the same regional distribution will vary slightly. Percentage coefficients of variation have been calculated, for all the different food expenditures, nutrient intakes, etc.

Procedure and Data Obtained. Each housewife who agrees to cooperate, is given a log book to complete for the succeeding seven days. For each day, she has to fill in four sections as follows:—

(1) Quantity, description and cost of each item of food bought on that day.
(2) Quantity, description, source of and price paid (if any) for all home-grown food, gifts and welfare food.
(3) Description of food served at each meal and note indicating which persons were present.
(4) Number and type of meals taken out by different members of the household.

In addition, the interviewer records on the cover page, details of family composition, including age, sex, relationship to housewife and occupation of all members of the household. This information makes the calculation of the nutritional needs of the family possible. Furthermore, food in the larder is weighed and measured at the beginning and end of the week, so that changes in stocks may be calculated.

Results. The results, calculated for the sample every month, show:

a. The average daily consumption per head of each food.
b. The average daily intake per head, and average daily requirement per head of each nutrient.
c. The average expenditure on and actual price paid for each food, and on all foods. Also the quantity and value of foods obtained from free sources.

(3) *The Ministry of Labour*

The Ministry of Labour has made less use of sampling than might perhaps have been expected. Reference has already been made to its 1937/8 Family Budget enquiry and to Professor Hilton's investigation. The other use of sampling which should be mentioned here is in the

Ministry's Analysis by Age-group of Insured Workers. Up to July 1948, at the time of the annual exchange of unemployment books, information about the distribution of insured workers by industry, sex and by four age-groups (under 16, 16 and 17, 18-20, 21 and over) was obtained. It was desired to obtain more detailed information for the 21 and over age-group and this was done first in 1937 and regularly since 1942 by means of sampling.

The sample was taken from bundles of about 100 unemployment books tied up at the Employment Exchanges (for dispatch to another office). Two books were selected out of every bundle, respectively $\frac{1}{3}$ and $\frac{2}{3}$ of the way through the bundle. The resultant 2 per cent sample (about 300,000) was a small one and the Ministry of Labour pointed out that great care needed to be taken in interpreting results.

These samples yielded analyses for total insured workers by sex, age and industry and by five-year age-groups. Up to 1947, the results are available only on a national basis; in that year, for the first time, a regional analysis was also published.

(4) *The Ministry of National Insurance*

The analysis based on the annual exchange of all unemployment books (not the additional sample analysis) not only yielded detailed information about the insured working population, but also served as a basis for estimates of the total working population.

With the new National Inusrance Act, which came into force on 5 July, 1948, a different type of insurance book is used and their large number (25,000,000) renders a simultaneous exchange of books impracticable.

The exchange of books is consequently being spread over the year at quarterly intervals and a careful sampling plan has had to be devised in order that detailed estimates for the whole working population may be made.

The books have been allocated to the four quarters in a systematic and unbiassed manner. No details have so far been announced about the kind of estimates it is intended to make at the quarterly dates, or the way in which they will be combined to give annual estimates.

(5) *The Ministry of Works*

The Ministry of Works uses two methods for obtaining details of the Labour Force in the Building Industry. In the first place, a quarterly census of the whole industry is taken yielding very detailed data on the distribution of the labour force by region, occupation, type of

work, etc. Owing to the large number of forms involved and the delay in getting them in, the first results are not available until 12 weeks after the Census date.

The other source of information is a monthly sample (started in 1945), for which the population consists only of the twelve main building trades (about $\frac{2}{3}$ of the whole industry). Information collected from the sample is confined to the total number of operatives employed by firms. Against these disadvantages of more limited coverage and less detailed information of the sample enquiry must be set the great economy for both the Ministry and the industry, increase in accuracy and the fact that the final tabulation is available three weeks after the sample date.

The Ministry of Works requires that the maximum error in the over-all total of operatives should be 1 per cent and found that the appropriate sample size would be between 5000 and 6000 contractors. The sample is based on the last available census tabulations (usually from a census taken 5–7 months previously). The sampling fraction varies from 1 in 100 for firms with 0 employees, 1 in 30 for firms with 1–5 employees to 1 in 2 for firms with 71–99 employees and a 100 per cent sample for all firms with 100 and more employees. The over-all sampling fraction is approximately 1 in 20, there being about 120,000 firms in the 12 Trades with which this sample is concerned.

The sampling in these investigations of the Ministry of Works is designed and executed with particular care and the possibility of taking a census less frequently and deriving more detailed information from the samples is being investigated.

(6) *The General Register Office*

The General Register Office is the department responsible for the Decennial Censuses of Population and for the collection and publication of demographic statistics generally. In view of the scope of its work, its failure so far to make use of sampling to any considerable extent is striking. There are two uses of sampling by the G.R.O. to which reference may be made:

(1) "Classification and tabulation of multiple or secondary causes of death" [8]. In the classification of death by cause, when more than one cause is mentioned, it is necessary to select one as that to which the death should be classed. Information should, however, be collected not only on the frequency of occurrence of each cause as a primary or secondary cause but also on the frequency with which certain causes appear in conjunction.

In each of the years 1921–1930, a group of causes of death was selected for further investigation. Apart from the ordinary punching of the primary cause of death on each card, the whole cause as certified was written at the top of the card. A sample was then taken for each group of deaths attributed to a particular primary cause; the sampling fraction varying from 1 in 10 for groups with a large number of deaths to a 100 per cent sample for groups with only small numbers.

The secondary causes of death on all the sampled cards were then coded and an analysis of associated causes was made.

(2) "Emergency Medical Service Records." Records are available for all patients who have been treated as in-patients under the E.M.S. A 1 in 5 sample of all the cards of discharged patients was taken, yielding about 45,000 cases for each year from the beginning of the war to the end of 1947. The data collected and tabulated from this survey included sex, age, civil status, branch of service or other occupation, and full details of the patient's hospital record, from his admission to his discharge.

It will be seen that both the above samples were samples from documents and the only field sample survey with which the General Register Office has, in fact, been in any way associated is the Survey of Sickness (mentioned in Section 1) the results of which are now published in the General Register Office Quarterly Returns.

Some general explanations of the apparent lack of enterprise regarding the use of sampling in Government departments will be offered in a later Section. It is, however, worth noting at this point that the Registrar-General is considering the use of sampling at the 1951 Census of Population but that no conclusions have so far been reached.

There is an unanswerable case for the use of sampling at the next Census to get out preliminary results more quickly. Furthermore there is no reason why all the final Census tabulations should be based on an anlysis of *all* the returns. Supposing that the same information was obtained from every member of the population (thus avoiding any legal difficulties), results of sufficient accuracy might be obtained on a number of questions by taking samples of the returns. This would save considerable time, labour and money and would avoid the situation arising out of the 1931 Census, some of the results of which it has still not been possible to analyse.

It is to be hoped that the Registrar-General will decide in favour of sampling and that no time will be lost in undertaking the necessary research and other preparations.

(7) *The Sample Family Census*

The Sample Family Census, which took place at the beginning of 1946, is an essential part of the work of the Royal Commission on Population. Its purpose was to obtain information about changes in the size of families and thus to provide data of obvious importance with regard to "the population problem and its bearing on housing, family allowances, social insurance, and other measures of social welfare" [9].

The sample of 1 in 10 of all the married women in the country (yielding about 1,600,000 women) was selected from the Food Office files. At the 1945 exchange of ration books, women had been asked to describe themselves as "Miss" or "Mrs." Accordingly, after every 10th reference page had been extracted from the file of persons aged 18 and over, the pages of all males and of all females describing themselves as "Miss" were discarded (apart from women who, though they described themselves as "Miss," had changed their names since the last exchange of ration books). The remainder then, were women who had described themselves as "Mrs." or who had not given the information. In order to preserve the randomness of the sample, *all* these women were contacted and those who were actually found to be "Miss" were then discarded from the sample. One more precaution was necessary before this could be regarded as the final sample of presently or formerly married women: the pages for women who had left the particular area were removed by checking the reference pages against the Maintenance Register. Great efforts were made to contact all the women sampled and substitutes were not permitted in any circumstances.

Questions were asked on:—whether at present married, widowed or divorced; date of birth and of first marriage (and of the end of the marriage, if applicable); dates of birth of all live-born children; number of children under 16 alive, and husband's occupation.

The information was collected by 12,000 enumerators who received 1/4d for each completed form. It is hoped to publish the report on the Census, with full details of the sampling methods employed and accuracy achieved, sometime next year.

(8) *The National Farm Survey*

The National Farm Survey was a development from the local farm surveys which were being carried out during the war to assist the work of the County War Agricultural Committees. The task of these Committees "may be shortly stated as ensuring that each farm makes its maximum contribution to food production" [10].

The National Farm Survey was carried out from 1941 to 1943 and consisted, in the main, of information obtained according to a *standardised* pattern from the local surveys and of the returns from the 1941 agricultural Census. The survey, in addition to aiding the local committees, provided invaluable statistical material on a national and county basis.

The Survey population consisted of the 300,000 holdings of 5 acres and over. Information was collected on the following subjects:—

A. Type of tenure, rent and length of occupation.
B. Economic type of occupier.
C. Size and type of holding.
D. Convenience of farm lay-out; situation of holding; condition of buildings.
E. Nature of the soil; nature of water and electricity supply.
F. Managerial efficiency of occupier and condition of cultivated land.

It was impracticable and unnecessary to base the national and county totals and averages on an analysis of the records of all the 300,000 holdings. A random sample was therefore drawn and was stratified by county and size of holding. The sampling fractions varied according to size of holding as follows:

Size of Holding Acres	Sampling Fraction %
5– 24.9	5
25– 99.9	10
100–299.9	25
300–699.9	50
700 and over	100

The final sample covered about 14 per cent of all the holdings in England and Wales. The figures throughout the report were arrived at by multiplying the sample results by the reciprocal of the respective sampling fractions. The full published report includes a detailed description of the sample design, the standard errors (which are trivial for national and very small for county data) and estimates of the gain through (a) stratification and (b) variable sampling fraction.

(9) *BBC Listener Research*

It is appropriate, at this stage, to say something about the work of the Listener Research Department of The British Broadcasting Corporation. This department, first set up in 1936, collects information

about the listening habits and tastes of the British public; it tries, in fact, to establish some kind of a "box-office" substitute in the world of radio. The head of the department has described the two principal tasks of Listener Research as being:—

(1) to find out how many people listen to each broadcast,
(2) to ascertain listeners' opinions of the broadcasts which they hear [11].

To accomplish these tasks, Listener Research organizes two separate and very different enquiries, which will now be described in a little detail.

The Survey of Listening

This continuous survey is based on a daily sample of 3000 persons (aged 16 and over). An equal number is interviewed in each of the six BBC broadcasting regions into which the country is divided, but the results are subsequently weighted according to the population in each region. A further sub-division into rural and urban areas is made within regions. Selection of individuals in each area is by means of quota sampling—according to quotas based on the sex, age and social class distribution of the population. The procedure employed for determining the social class of the informant is for the interviewer to ascertain his occupation and with the aid of a "common-sense assessment of the contact's manner, bearing and conversation" to classify him as working, lower middle or upper middle class. It may be questioned whether this procedure is reliable and, considering the relevance of social class in this type of survey, one might suggest the use of a more precise method.

In the interview, the contact is asked to name the programmes he listened to on the previous day. The interviewer will try to aid the informant in remembering the broadcasts and may have to distinguish between "listening to" and simply "hearing" a programme. This, of course, is the great snag of all radio research and it is doubtful whether there is any satisfactory solution.

The results of these daily interviews are issued about 7 days after the broadcast in the form of what is called the Listening Barometer; that is to say, against each programme are shown the percentages of the sample (in each region and in Great Britain as a whole) who listened to it.

The Listening Panel

While it is of great value to the BBC to know the number of people who listened to each broadcast, it is equally important to know how

a particular broadcast was liked by those who did actually listen to it. For this purpose, the BBC has established a panel made up of volunteers recruited as a result of occasional appeals. This panel consists of 3600 members, distributed equally over the six regions. Each panel member receives, twice a week, a questionnaire containing questions about a number of broadcasts to be made during the next few days. All types of programmes are included and panel members are specifically asked to give their opinions only on those broadcasts in which they were anyhow interested; that is to say, duty listening is to be avoided. The panel members (whose average "panel life" is 18–24 months) are, of course, self-selected and it is questionable whether the opinions sent in on a particular broadcast can be regarded as fully representative of the opinions of all who listened to it. The summarised results of the questionnaires are passed on to the producers of the programmes.

The results of Listener Research are given the widest possible circulation inside the BBC, but it is impossible to assess the extent to which they influence programme policy and planning. As Listener Research constitutes the main link between the BBC and its public, directors of sections and programme producers probably take a considerable interest in its results.

The above account of the use of sampling in official surveys cannot be regarded as complete. It has not been possible, for instance, to say anything about the use of sampling in the Ministry of Home Security's wartime investigations of the effects of bombing; or of the sample enquiry which the Royal Commission on the Press is believed to be making; or for that matter, of the employment of sampling methods in the internal work of Government departments. It is believed, however, that enough has been said about the most important official sample surveys, for which any details could be obtained, to form a basis for the general remarks which follow in the last sections of this paper.

B

SURVEYS CONDUCTED BY SEMI-PUBLIC BODIES AND RESEARCH INSTITUTES

An indication has already been given of the wide use of sampling in social surveys and there is an increasing tendency, on the part of Universities, Research Institutes and local authorities, to avail themselves of sampling techniques in their surveys of social and economic problems. It is now proposed to describe a few of the more important and interesting surveys which have recently been, or are still being, carried out in this field by non-official, non-commercial organizations.

UNIVERSITY OF MICHIGAN LIBRARIES

1. *Population Investigation Committee—Survey of Maternity*

The survey on maternity [12] was undertaken in order to obtain data on the social and economic aspects of childbearing. In particular, information was sought on the availability and use made of maternity services in different parts of the country and social classes, and on present-day expenditure on childbirth.

It was desired to collect the information through personal interviews with mothers "whose experience could be regarded as typical of the whole population of women now bearing children." The method adopted was to take a sample in time *from all* local authorities—rather than a sample *of* local authorities—and to interview *all* women who were delivered in England, Wales and Scotland during the week of 3–9 March, 1946. This procedure had the advantage of greatly easing and speeding up both the field work and the administration of the survey.

Officials of all the 458 local authorities were approached and active support was received from 424 (92%) of them. It is estimated that 16,695 births were registered during the week in question, of which 15,130 were notified to the Survey Committee. Successful interviews were made regarding 13,687 (or 90.5%) of these. It is interesting to note that refusals were obtained in only 2% of the cases.

As a great deal of information had to be asked for, it was decided to issue two separate types of questionnaires. Type 'A' dealt mainly with the use made of maternity services, while Type 'B' concentrated on the financial aspects. Certain basic questions were, of course, asked in both. The 424 authorities were split at random into two groups, one group for the Type 'A' and the other for the Type 'B' questionnaire. The proportion of successful interviews was 91.5% for Type 'A' and 89.3% for Type 'B' questionnaire. Analysis shows that, in most important respects, the samples of mothers making up the two surveys are closely comparable.

The results of the questionnaire inquiry—which, together with the methods, are very fully described in the report—are supplemented by detailed studies of the extent and quality of maternity services in selected areas.

One aspect of particular interest remains to be noted. It was decided to conduct a follow-up survey in order to study

(a) differences in morbidity of different groups of mothers and children;
(b) the factors which influence the health and development of different groups of children during their first two years of life;
(c) differences in morbidity and development of full-time and premature children.

It was felt to be unnecessary to apply the questionnaire to all the mothers in the original enquiry, as long as the absolute number in each social group was large enough to yield the precision required. For (a) and (b), the sample consisted of one out of four mothers and children in the "manual workers" group (chosen at random) and all those in the other groups. This gave a total of about 5000 cases.

This follow-up survey took place in March 1948 and met with considerable success. Special efforts were made to trace those women who had left their original address and completed questionnaires have been received from about 94% of the women.

For the survey concerning premature children, a different method of selection was employed. Each of the 800 or so premature babies in the original sample was matched with a full-time one of the same sex, social class and birth order and born to a mother of the same age, living in the same region and at the same room density. A number of the babies subsequently died while others moved out of the original area, so that a good number of the pairs had to be abandoned. It has, however, proved possible to follow-up 640 pairs through their first two years of life. This unusually large-scale and carefully conducted controlled investigation will, in fact, provide invaluable material on the causes of prematurity; on possible ways of preventing deaths arising from it; and on the question as to the extent to which premature and full-time babies differ in their subsequent development.

2. *Population Investigation Committee—Inquiries into the Trend of Intelligence*

In 1947, the Population Investigation Committee and the Scottish Council for Research in Education jointly initiated an inquiry into changes in national intelligence in Scotland. The purpose of the inquiry is to test the hypothesis that the "current patterns of differential fertility (which show a negative correlation between size of family and the measured intelligence of the children) really imply a fall in national intelligence."

The present inquiry has been conducted in Scotland, because in 1932 a complete age-group of Scottish children had been the subject of an inquiry into intelligence and it seemed appropriate to apply the same tests, which had then been used, under similar conditions to the same *age-group* of children. The survey has covered some 80,000 children born in 1936, in respect of whom the same group intelligence test has been administered and for whom an individual questionnaire was completed, giving details of age, sex, size of and position in family, school and class, etc.

UNIVERSITY OF MICHIGAN LIBRARIES

A special and more detailed questionnaire has been filled in by a random sample of children, consisting of all children born on the first three days of each month in 1936 ("36-day sample"). Furthermore, individual Binet tests have been administered to a further sub-sample consisting of all children born on the first day of each alternate month ("6-day sample").

The field work has been highly successful. The group tests were administered to 88% of the children and a greater number completed the questionnaire; the more detailed questionnaire was obtained from 99% of the 36-day sample, while the Binet tests were given to 99% of the 6-day sample.

Results, which are not yet available, will show the change in the average and standard deviation of the I.Q. since 1932; and the relationship between I.Q. and size of family, parental occupation and other factors. It is intended to follow up a sample of the children, in order to compare the development of children of different I.Q.'s over a number of years. For this purpose, the 6-day sample will be taken plus 400 children of very high I.Q. and 400 children of very low I.Q.

3. *London School of Economics Social Research Division—Survey on Social Class and Social Mobility*

The aim of this survey, financed by the Nuffield Trust, is to discover what people mean by social class, and what are the chief factors accounting for class differences and for the movement from one class to another. An attempt will be made to estimate the relative importance of these factors and to trace the changes which have taken place in their influence over past years.

A start has been made with a factual survey carried out in two London Boroughs and a neighbouring rural area, designed to test a method of obtaining information about educational opportunity and occupations entered in successive generations. The possibility of extending this investigation to other parts of the country is under consideration.

The aim in this survey was to interview a married person, preferably the wife, and the introduction to the married person was obtained by drawing a random sample of individuals from all those of age 16 and over resident in the area investigated. If the individual drawn was unmarried, but living with parents, the mother was to be interviewed. If the individual drawn was of age 60 or over living with married son or daughter, the young wife was to be interviewed. Persons not to be interviewed were:

(a) Unmarried persons living apart from their parents.
(b) Persons previously married (to avoid complications possibly arising from differential treatment of step-children).
(c) Separated or widowed persons.

It is expected that this research will extend over approximately five years.

4. *Local Social Surveys*

The early social surveys, to some of which reference has already been made, were primarily focused on the central problem of poverty and its different aspects. In the late '30's, the emphasis began to shift. Attention was being directed more and more to questions connected with town planning and life on housing estates as, for instance, in the surveys of Birmingham [13], Becontree [14] and Watling [15]. Since then, and especially since the end of the war, a large number of regional planning surveys have been initiated and published. The social survey has rightly become an integral part of town planning and most of the recent planning studies will be found to contain a social and economic survey, usually based on a sample of dwellings or households. Thus, for instance, the Middlesbrough Plan [16] includes a sample survey as do the plans of Luton [17], Worcester [18] and some of the studies made by the Association for Planning and Regional Reconstruction. The sampling methods employed are generally not of great intrinsic interest and it is proposed here to describe, as an example, only the survey of Luton, which is one of the best published so far.

Report on Luton (1945)

Luton, described as a "young, expanding industrial town" surrounded by rural areas, lies 20 miles from the northern fringe of London and has approximately 100,000 inhabitants. It was felt that, in view of current housing problems and the present-day changes in education, health and other services, the Local Council should have up-to-date information on population, housing and the social aspects of public services.

As the authors were primarily interested in population and housing, they would have liked to survey every occupied dwelling. This, however was impossible and a sample of dwellings was taken. Luton was divided into three areas. District I—areas of unfit houses; District II—the older parts of Luton; District III—the rest of the Borough. The population density of these districts is shown by the average number of

UNIVERSITY OF MICHIGAN LIBRARIES

persons per acre, which were 92, 36 and 11 respectively. The sampling fraction varied as follows:

District	No. of Houses	Sampling Fraction	No. of Houses Surveyed
I	890	1/1	890
II	5,270	1/5	1,054
III	22,300	1/10	2,230
Total	28,460		4,174

To obtain data for the whole Borough, a sample of 1 in 10 of all dwellings in the borough was achieved by using 1 in 10 of the houses sampled in District I, every other one in District II and every one in District III. The actual addresses in each district were selected from the Rating List.

The survey provided data of three kinds:—a certain amount of information for each sampled household (e.g. no. of rooms, no. of persons, overcrowding); information for each individual (e.g. age, sex, civil status, occupation, education); and detailed fertility data relating to each married woman. On the basis of this full and carefully obtained information, the authors were able to give invaluable guidance for future planning.

C

PUBLIC OPINION AND MARKET RESEARCH

It is not intended in this article to give an exhaustive account of the surveys conducted or methods employed in opinion polls and market research in this country. Some of the main organisations in the field will be mentioned and an indication will be given of any interesting characteristics they may possess. The justification for this somewhat sketchy treatment is twofold. In the first place, it is in the nature of the work of, at any rate, the market research bodies that the aims, methods and results of their surveys are made known to their clients rather than to the general public. Anything like an accurate account of their activities would be very hard to give. In the second place, it is very much doubted whether either the methods employed by public opinion polls and market research bodies here or the fields of application to which they direct their attention differ sufficiently from American practice in this field to warrant a full discussion in this article.

(i) *British Institute of Public Opinion*

The B.I.P.O., founded in 1936, is one of the international chain of

Gallup institutes. Opinion polls in this country do not, of course, enjoy anything like the publicity given to polls in America. Thus, the findings of the B.I.P.O. are published by only one newspaper, the daily News Chronicle. Nor could it be said that opinion polls here yet arouse such universal interest or such bitter controversy as in the United States.

The B.I.P.O. has, in the past 12 years, conducted polls on a wide variety of topics and has also published fairly accurate forecasts on the 1945 general election and on numerous bye-elections. Apart from the work concerned purely with opinion polls, B.I.P.O. has, on a number of occasions, co-operated in, or conducted surveys for, other organisations. Thus, it undertook part of the war-time body-weight survey for the Ministry of Food and several surveys for the Board of Trade. The method of sampling used by B.I.P.O. is quota sampling.

(ii) *Mass Observation*

The surveys conducted by Mass Observation are not based on scientific sampling methods. Mass Observation is mentioned here because it has received very wide publicity and aroused considerable interest. Founded in 1937 it has, in a very active decade, published some 14 books and several hundred bulletins and articles.

Its reports are based on two different sources: field surveys and a national panel of respondents, neither of which appear to be based on a randomly selected sample.

(iii) *Market Research Bodies*

As in America, there is in England a large and increasing number of market research organisations. As far as can be judged, from the few leaflets which it has been possible to obtain, all of these organisations claim that scientific sampling methods form the basis of their results and standard error formulae are usually quoted. Not enough is published to permit any evaluation of the methods employed by market research bodies; it can be stated, however, that the method of selection generally employed is quota sampling which is cheap (judged purely in money terms), offers great ease of administration and fieldwork (it by-passes the problem of call-backs), and is, for commercial purposes, considered as sufficiently accurate.

The most important organisations in the field are probably the British Market Research Bureau and Research Services Ltd. The former, among other activities, has done a number of surveys for the Board of Trade and the Ministry of Food. Research Services Ltd., is the new name of what was formerly the Research Department of the London

UNIVERSITY OF MICHIGAN LIBRARIES

Press Exchange. It now constitutes a separate company and covers a wider range of activities than other market research bodies. Another company, Attwood (Statistics) Ltd., runs a Consumer Panel, selected on the basis of a random sample.

D

A COMPARISON BETWEEN BRITISH AND AMERICAN SAMPLING PRACTICE

The preceding sections show that investigations on a wide variety of subjects are undertaken by means of sampling, both inside Government departments and by other organisations. Nevertheless, it can hardly be claimed that the available sampling techniques are exploited to the full or that sufficient effort is made to experiment with different and more refined methods. The remainder of this article is devoted to a discussion—based on the foregoing survey—of the general features distinguishing the use of sampling in Great Britain.

It is important, in this discussion, to guard against the error of treating the American and British situation as strictly comparable. The enormous area and widely dispersed population of America gives rise, in the first place, to a much greater *necessity* for sampling and has led to the development of sampling methods particularly suited to these circumstances. In Great Britain, the population is relatively small and highly concentrated. It would be a mistake to believe that American sampling methods could necessarily be used with the same success here, and much experiment and research is needed to decide which refinements and developments can usefully be "imported" (it is interesting to compare a French point of view on this subject [19]).

Further to this, it should be pointed out that probably, in this country relatively more data about the population comes in as a by-product of administration and does not need to be obtained by special surveys. (Information about the Labour Force is a case in point.) This difference again implies that the *necessity* for sampling is not usually as urgent as in the United States, but it also points to far greater *opportunities* in the sampling of documents and returns to obtain quick and regular data to aid administration. It cannot be said that anything like full use is being made of these opportunities.

(a) *The Attitude to Sampling in Great Britain*

It must, in the first place, be emphasised that in England sampling is not yet as *generally* understood or accepted as a reliable way of collecting information as in the United States. The American public

was made "sample conscious" by the evident success and through the enormous publicity of public opinion polls in general and their past election forecasts in particular. The corresponding absence of such publicity here partly explains any lack of confidence in sampling felt by the general public.

More important from the point of view of the development of sound sampling practice is the distrust and lack of knowledge remaining, in Government departments and elsewhere, among persons who are potential initiators of sample surveys. No doubt, now that more attention is paid to sampling in University statistics courses, this distrust and ignorance will gradually be dispelled. Yet, there is an urgent need for the spread of knowledge through the full publication of the methods and results of surveys and for a regular procedure by which statisticians and administrators in Government departments and elsewhere may avail themselves of advice on the conduct of sample surveys.

A department, perhaps within the Central Statistical Office, should devote its whole time to the examination of past and current sample surveys in all fields. By scrutinising the methods used in the past and by giving advice on new projects, such a department could help first to establish and ultimately to maintain a high standard of sampling practice. Further, it could act as a central Register of work being done by all sorts of organisations and thus prevent much duplication of research.

(b) *The Publication of Sample Surveys*

It is obvious that the full publication of survey methods and results is a considerable asset to further development. In England, unfortunately, the position regarding publication is very unsatisfactory.

Some indication of the circulation of the Social Survey reports was given earlier; the Family Food Survey of the Ministry of Food, which has continued for seven years and is the main source of information concerning the diet of the British public, is only now receiving publication; as far as can be ascertained, the sample surveys undertaken by the Ministry of Works, the Board of Trade, the Ministry of Labour and the Ministry of Home Security have not been described in any publication.

Non-publication of sampling methods is only one aspect of a general, but possibly decreasing, tendency in Government departments to treat their information as confidential or, at least, not for publication. This is a great barrier to the development of sampling methods and the spread of knowledge. It is to be hoped that the valuable suggestions concerning this problem, made in a recent report by the National

UNIVERSITY OF MICHIGAN

Institute of Economic and Social Research [20], will be followed up. In all fairness, it must be added that there are exceptions to this tendency. Thus, the National Farm Survey was fully described in the official and publicly available report. Similarly, the methods as well as the results of the Sample Family Census will be given in great detail in the forthcoming report on that survey.

There is one more point on this question of publication. With the increasing use of sampling in investigations concerned with human populations, there appears to be a good case for an international journal devoted to this subject. Such a journal would not need to confine itself to sampling, but could range over all the methodological problems which arise in the planning and execution of surveys. Perhaps the United Nations Sub-Commission on Statistical Sampling might consider the question.

(c) *Methodological Research*

The work on sample surveys in this country is largely confined to *ad hoc* surveys or, at any rate, surveys which are important for their results rather than their methods. There is clearly a need for more systematic and coordinated experimentation and research on sampling techniques, on the lines of the work done, for instance, by the Bureau of the Census. Perhaps this task could most suitably be performed by a department such as that suggested in (a) above.

(d) *Area Sampling and Quasi-Random Sampling*

This is not the place to enter into the controversy regarding the respective merits of these two methods of sampling. It is desired merely to indicate why area sampling, which has been so successfully applied in the United States, is not used here; and whether it is likely to be used in the future.

The most obvious reason for the non-use of area sampling here lies in the availability of the National Lists, mentioned earlier. None of these lists is perfectly accurate, but if care is taken, they can form an adequate basis for random sampling at given intervals, i.e., quasi-random sampling. The fact that these lists are available, while maps suitable for area sampling are not (the maps of the Fire Service may be closest to what is required) has acted as a check to experimentation with area sampling methods.

It is probable that even with the relatively small and concentrated population of Great Britain, area sampling would lead to certain

advantages. Research is needed to show whether the high initial outlay would be justified by increased accuracy and economy resulting from its use.

(e) *The Use of Purposive Selection*

It has been seen that, in Great Britain, most of the sample investigations concerned with human populations rely on stratified random sampling from the lists as their method of selection. If national estimates are involved, the sample may have to be spread geographically over the whole country. Often, cost considerations and interviewer resources make it desirable to get the field work reasonably concentrated and, in a number of surveys, the primary sampling units are chosen by purposive selection. Thus, in the Ministry of Food Family Diet Survey, the towns are chosen to be representative of Great Britain, with regard to their size and the character of the region. The same sort of selection is used in some of the Social Survey investigations.

Selection of the primary sampling units on the basis that they are "representative" may easily introduce bias and it will be better to use the alternative procedure of sampling at random from a list of areas or towns. The areas would be grouped in strata in some appropriate way and a variable sampling fraction could be applied, perhaps on the lines used in the United States Sample of the Labour Force [21].

(f) *The Use of Quota Sampling*

Within the primary sampling units, the usual procedure employed for the selection of the final sampling units is quasi-random sampling. It should, however, be noted that quota sampling is still the method generally used by commercial organisations, opinion research bodies (including Listener Research) and occasionally in Social Survey investigations.

The cheapness and ease of execution of quota sampling are qualities which endear it to commercial organisations. But, from a statistical point of view, the method is unsatisfactory for three reasons. In the first place, the choice of units depends on human judgement and, therefore, may involve serious bias. In the second place, quota sampling does not permit any effective control over the work of the interviewers; and finally, it is not possible to attach an estimate of the sampling error to the results. The method should, therefore, not be employed in any surveys on which administrative decisions are to be based.

UNIVERSITY OF MICHIGAN LIBRARIES

(g) *Further Uses of Sampling*

There are a number of fields to which sampling might well begin to be applied in this country and there are some indications already that new ground is being broken and that further uses are being contemplated.

The Board of Trade has, at its disposal, a fairly complete list of industrial firms in the country. This might serve as the basis for samples designed to obtain information additional to that supplied by the Censuses of Production and Distribution. It is believed that the lists of firms in specified manufacturing industries are to be utilized for sample surveys on Capital Expenditure.

Much information on family expenditure is needed and will regularly have to be obtained for any contemplated permanent Retail Price Index to take the place of the present interim Index. Sample surveys have been started by the Social Survey with a view to obtaining information on the national expenditure on selected items. So far, surveys on laundry and household textiles have been conducted and ultimately all types of consumer expenditure will be covered.

The most obvious gap in sampling practice in this country is in the field of population. It is to be hoped that the Registrar-General will decide to use sampling in the 1951 Census and that he will also begin to consider the possibility of taking regular sample surveys of the population to fill in the 10-year gaps between Censuses. Sampling for population estimates requires more refined techniques than are customary in most of the surveys mentioned above. It is in this field that American developments have been most striking and any contemplated use of sampling at the next Census of England and Wales should be based on a very careful examination of American experiences.

The use of sampling can lead to great economy in cost, time and personnel and is bound to play an increasingly important role in Government administration and elsewhere. The most urgent needs in this country are for systematic research, in government departments, universities and other bodies, into sampling as well as other (and often more difficult) problems connected with surveys; and for full publication of the methods used in and lessons learnt from completed sample surveys.

REFERENCES

[1] Report of the Sub-Commission on Statistical Sampling (United Nations E/Cn. 3/37) issued 21st October, 1947.

W. E. Deming "A Brief Statement on the Uses of Sampling in Censuses of Population, Agriculture, Public Health and Commerce" published by the United Nations, Lake Success, 1948.

Series of reports submitted to the 2nd meeting of the Sub-Commission at Geneva in September 1948 (United Nations E/Cn. 3/sub. 1/various).

[2] Bowley, A. L., "Working-class Households in Reading," *Journal of the Royal Statistical Society*, 1913.

[3] Hilton, John, "Enquiry by Sample: an Experiment and its Results," *Journal of the Royal Statistical Society*, 1924.

[4] Stephan, F. F., "History of the Uses of Modern Sampling Procedures," Presented at the 25th Session of the International Statistical Institute and reprinted in the *Journal of the American Statistical Association*, Vol. 43, March, 1948.

[5] Ministry of Labour, "Weekly Expenditure of Working-class Households in the U.K., 1937–8," *Ministry of Labour Gazette*, Dec. 1940, Jan. and Feb. 1941.

[6] *Interim Report of the Cost of Living Advisory Committee.* H.M.S.O. Cmd. 7077; and "Index of Retail Prices" *Industrial Relations Handbook 1944*, Supplement No. 2, January, 1948.

[7] Box, K. and Thomas G., "The Wartime Social Survey," *Journal of the Royal Statistical Society*, 1944.

[8] *Registrar-General's Decennial Supplement for England and Wales*, 1931, Part IV (published 1947).

[9] Article on the Sample Family Census in the "*Times.*" January 17th, 1946.

[10] Ministry of Agriculture, "*National Farm Survey of England and Wales*," A Summary Report. H.M.S.O. 1946.

[11] Silvey, R. J. E., "Methods of Listener Research employed by the B.B.C.," *Journal of the Royal Statistical Society*, 1944.

[12] Population Investigation Committee, "*Maternity in Great Britain*," Oxford University Press, 1948.

[13] Bourneville Village Trust, "*When We Build Again*," 1941.

[14] Young, T., "*Becontree and Dagenham*," 1934.

[15] Durant, Ruth, "*Watling*," 1939.

[16] Lock, Max and others, "*The County Borough of Middlesbrough Survey and Plan*," 1947.

[17] Grundy, F. and Titmuss, R. M., "*Report on Luton*," 1945.

[18] Glaisyer, Janet and others, "*County Town*," 1946.

[19] Thionet, P., "*Méthodes Statistiques Modernes des Administrations Fédérales aux Etats-Unis*," Paris, 1946.

[20] National Institute of Economic and Social Research, "*Social and Economic Research and Government Departments*," 1947.

[21] Hansen, M. H. and Hurwitz, W. N., "*A New Sample of the Population*," Bureau of the Census, 1944.

UNIVERSITY OF MICHIGAN LIBRARIES

UNEMPLOYMENT AND MIGRATION IN THE DEPRESSION (1930–1935)*

RONALD FREEDMAN
AND
AMOS H. HAWLEY
University of Michigan

This is a study of the reciprocal relationship between migration and unemployment in Michigan during the depression period 1930 to 1935. Specifically, it is concerned with two questions:

(1) Do the migrants during a depression have a poor employment history as compared with non-migrants of similar characteristics?

(2) Do the migrants have a better employment experience after migration than non-migrants of similar characteristics?

By the use of matched control groups it is found that the differential in unemployment rates occurs after migration, not before. The results are consistent with the hypothesis that in a depression migrants tend to be at a disadvantage in the new labor market to which they move.

THE PROBLEM

THE QUESTIONS above are at the heart of the more general problem of the relationship between migration and economic opportunity. Comparisons of the pre-migration employment status of the migrants with similar persons who stay at home should indicate whether unemployment is the important "push" factor in migration it is frequently believed to be. Comparisons of the post-migration employment history of the migrants with those of non-migrants at the source-point should help to indicate whether migrants succeeded in improving their employment status by making their move. Similar comparisons with the non-migrants at the destination point can help to indicate whether the migrants are at a disadvantage in competing for economic opportunities with the resident population.

Most studies of these questions have failed to give convincing answers even for a specific time and place. Some studies have failed, because they have focused on "distress" migrants without considering the relative importance of such migrants in the total stream of migration. The result has been a stereotype of the migrant in a role such as that of

* This study was made possible by a grant from the Faculty Research Funds of the Horace Rackham School of Graduate Studies of the University of Michigan.

the "Okie". Other studies have failed to compare the migrants with similar non-migrants at either end of the migration process, so that it is difficult to determine whether the employment history of the migrants was distinctive and whether it was a product of migration or of some other factor.

There is an attempt in the present study to minimize the first difficulty by considering a sample of migrants in the area studied without restriction as to their "distress" position. The second difficulty is minimized by comparing the migrants with carefully selected "control" groups of non-migrants at each end of the migration process. An important methodological objective of this study is to experiment with the use of the matched-group control method in the population-migration field. The authors recognize that the data of this study are limited to a specific time period and to specific streams of migration. Many further studies are needed to permit valid generalizations. It is desirable that such studies should refer to migration between specific areas and should control as many relevant factors as possible.

THE DATA AND METHODS

The data for this study are from the Michigan Population and Unemployment Census of 1935.[1] The distinctive feature of the schedule for this census is a month-by-month history of the occupation and employment status of every person 15 years old or over for the period April, 1930, to January, 1935. This history includes data on place of employment, which makes it possible to identify migrants and their routes of movement.

The sample for the present study consists of all those white male migrants to Flint or Grand Rapids from other places in Michigan, who were at least 25 years old at the time of migration to the cities. Males 15 to 24 years old were excluded from the study in order to deal only with those migrants whose eductaion was presumably completed. The female migrants were reserved for a later study, because their employment status involves special factors. Non-white migrants, originally included in the study were eliminated, because they were found to constitute a negligible proportion of the intra-Michigan migrants to Flint and Grand Rapids. Flint and Grand Rapids were selected as the destination points to be studied, because they are very different both

[1] *Michigan Census of Population and Unemployment*, Michigan State Emergency Welfare Relief Commission: Lansing, 1937, Nos. 1–10. The sampling procedure for this study involved complete enumeration of all cities in Micigan between 3,000 and 40,000 in population, a 20 per cent sample in each city of 40,000 and over, and a complete enumeration of a random selection of 20 per cent of all the towns and villages (under 3,000) and of open country type rural townships.

UNIVERSITY OF MICHIGAN LIBRARIES

with respect to population history and economic base. Therefore, it was hoped that studies of the migrants to each place might be treated as separate "experiments." Agreement of the findings for the two cities should give them greater validity. The study was necessarily limited to intra-Michigan migration, because data were not available for non-migrants outside of Michigan. After observing the limitations noted above in selection of the sample, the sample of migrants to Flint was found to number 360. The number in the Grand Rapids sample was 186.

These samples are the basis for a series of studies of the relationship between migration status and a number of other characteristics (education, occupation, occupational mobility). This paper on the relationship of migration and unemployment is only one part of the larger series.

For the study of unemployment, each migrant was "matched" with a "control" non-migrant at the place from which he came and another at the place to which he moved (either Grand Rapids or Flint). The characteristics used for matching were age (within 3 years), occupation (in terms of the major census socio-economic groups), occupational history, (in terms of change between socio-economic classes), education (within 2 years of school achievement), and marital status. For every characteristic, except marital status, the matching was done as of the date of migration. Data on marital status were available only as of the end of the period. Nevertheless it was used as a basis of matching, because it was felt that the error involved would be considerably less than that involved in omitting this factor as a control. In a very few cases migrants in the study had come from places of less than 3,000 population which had not been included in the census sample. Since it was impossible to match such migrants at the actual source-point, they were matched with non-migrants from a similar place located near the actual source-point. Three of the cases of failure to match migrants are accounted for by the failure to find a substitute source-point. The total number of such successfully made substitutions is 9.

Although the choice among "eligible" matches was not made in the most rigorous manner possible, it is believed that the procedure used did not introduce any systematic biases. The procedure was essentially to enter the schedules at a point determined by a system of random numbers, examine the schedules serially until a match was found, then to select a new starting point and to repeat the process. Several alternative methods involving selection of all possible eligible matches and random selection among them were abandoned as prohibitively time-consuming. There does not seem to be any reason to believe that systematic errors were introduced by the method used.

Although it was not possible to match all the migrants at both ends of the migration, proportions matched are large enough so that the authors are confident that the results of the study are not affected by the unmatched cases. Of the 360 migrants to Flint 312, or 87 per cent were matched at the destination and 296, or 82 per cent, at the source-points. Of the 186 migrants to Grand Rapids 171, or 92 per cent were matched at the destination and 149, or 80 per cent at the source-points. This represents a considerably higher proportion of matched cases than are found in most comparable "matched-group" studies. A study of the unmatched migrants indicates that their employment history is not sharply different from that of the matched cases.

Several specific problems in connection with the matching need to be explained. First, those migrants whose occupational classification was "farmer" or "farm-laborer" obviously could not be matched identically for occupation at the destination point. The procedure followed was to match farmers with members of the "proprietors, managers, and officials," group and to match "farm-laborers" with "common laborers." Although these are the groups closest to the farm occupational categories in the census hierarchy of socio-economic classes, the matching on this basis is admittedly makeshift. Therefore, in the specific comparisons of migrants and non-migrants at the destination the effect of including or excluding farm migrants has been evaluated in each case. In any event, the matching permits a comparison of the farm migrants with the urban occupational group with which they are most frequently contrasted and even combined.[2]

In the study of unemployment a second problem arose concerning those persons who were either unemployed or out of the labor force during the entire period from 1930 to date of migration. It was not possible to classify these persons by occupation during the pre-migration period. In each of these cases the match was made as if "unemployment" or "out of the labor force" were the pre-migration occupation, respectively. This means that the employment status of these two categories of migrants was artificially perfectly matched for the pre-migration period. These cases can be compared, however, for the post-migration period for which there were no artificial restraints. During the pre-migration period the comparisons are valid only for those who were employed at some time. In those cases where either a migrant or a control non-migrant was found to be out of the labor force during part of a period, this part of the period was omitted from the comparison

[2] In a number of tabulations of the 1940 U. S. Census for cities farm operators were tabulated with "proprietors, managers, and officials," and farm laborers were tabulated with "laborers."

UNIVERSITY OF MICHIGAN LIBRARIES

for the pair. The periods excluded for all groups taken together were relatively small.

We matched 271 of the Flint migrants and 140 of the Grand Rapids migrants at both ends of the migration process. Comparison of tables based on these completely matched (both end) cases with the tables based on migrants matched at only one end of the comparisons have resulted in no essential differences. Therefore, in this paper the tables and discussions are based on the larger number of cases, including one-end matches.

On the average the migrants made their moves about 37 months after the beginning of the period, so the average pre-migration period is about 3 years. One important limitation of the data for this study needs to be mentioned. Data are available only for "resultant" migrants, that is those who moved to Flint or Grand Rapids between 1930 and 1935 and were still there in 1935. Data for the migrants who moved again away from Flint or Grand Rapids before 1935 are not available.

THE FINDINGS

In the pre-migration period, the migrants to Flint or Grand Rapids had a somewhat higher rate of unemployment[3] than the non-migrant control group at the source-points. However, in neither case was the unemployment rate of the migrants or the differential between migrants and non-migrants large enough to justify considering the personal experience of unemployment as the important causal factor in migration.

The data on pre-migration employment status are found in Table 1 for the migrants and their source matches. In comparing migrants and non-migrants during the pre-migration period those unemployed during the entire pre-migration period should be excluded, since these migrants and non-migrants are artificially equated on unemployment. Excluding this group, 18 per cent of the Flint migrants and 13 per cent of the non-migrant source matches were unemployed in the pre-migration period. For Grand Rapids the corresponding figures are 13 per cent and 11 per cent. Even if those totally unemployed during the pre-migration period are included in these rates only 24 per cent of the Flint migrants and 21 per cent of the migrants to Grand Rapids were unemployed in the pre-migration period. The migrants do not appear to be a group marked by an unusually high rate of unemployment for the period in question. In this and other comparisons to follow, differ-

[3] The "unemployment rate" used in this study refers to the percentage of persons in any group who were unemployed at any time during the period in question.

ences will appear as between unemployment rates for Flint and Grand Rapids for each type of migrant or non-migrant. Such differences are to be expected on the basis of the different economic structures of the two cities. However, for the two cities the differences between the migrant and non-migrant unemployment rates are in the same direction in almost every comparison. Variations in the absolute rates of unemployment as between the two cities are not relevant to the problems being investigated.

TABLE 1

PERCENTAGE DISTRIBUTION OF EMPLOYMENT STATUS OF MIGRANTS AND NON-MIGRANTS MATCHED AT SOURCE, FOR PRE-MIGRATION PERIOD FLINT AND GRAND RAPIDS

Employment Status in Pre-migration Period	Migrant Status and 1935 Residence			
	Flint		Grand Rapids	
	Migrants	Non-migrants	Migrants	Non-migrants
Total percentage:	100.0	100.0	100.0	100.0
In labor force:	99.0	99.0	99.3	99.3
Always employed	76.0	81.1	85.2	87.2
Always unemployed	5.4	5.4	0.7	0.7
Sometimes unemployed	17.6	12.5	13.4	11.4
Out of labor force	1.0	1.0	0.7	0.7
Number	296	296	149	149

A check on the relative numerical importance of the group of migrants unemployed throughout the pre-migration period was provided by the selection of "random" samples of non-migrants at source and destination points for each group of migrants. The only restriction on selection was that the non-migrants should be at least 25, male, and white. Each migrant group thus had a random non-migrant "comparison" group. These non-migrants were paired at random with migrants to determine a date for separation of the "pre-migration" and "post-migration" periods. At both ends of the migration process and in both cities the number of migrants unemployed throughout the pre-migration period was approximately equal to the number so unemployed among the random comparison groups. Whether considered on an absolute basis or relative to non-migrant comparison groups those unemployed throughout the pre-migration period were not an important group.

It is interesting to not that in the pre-migration period the employment status of the migrants did not compare unfavorably with that of

non-migrant controls at the destination points. As Table 2 indicates, there was less than 1 percentage point difference in the unemployment rates of migrants to Flint and the non-migrant controls in Flint; 19.9 per cent of the migrants and 20.5 per cent of the non-migrants were unemployed in this comparison. The unemployment rate among migrants to Grand Rapids was even less than that of non-migrants in Grand Rapids in the pre-migration period. The unemployment rates were 15.2 and 21.1 for migrants and non-migrants respectively.

TABLE 2

PERCENTAGE DISTRIBUTION OF EMPLOYMENT STATUS OF MIGRANTS AND NON-MIGRANTS MATCHED AT DESTINATION FOR PRE-MIGRATION PERIOD FLINT AND GRAND RAPIDS

Employment Status in Pre-migration Period	Migrant Status and 1935 Residence			
	Flint		Grand Rapids	
	Migrants	Non-migrants	Migrants	Non-migrants
Total percentage	100.0	100.0	100.0	100.0
In labor force:	98.4	98.4	98.9	98.9
Always employed	71.8	71.2	81.9	76.0
Always unemployed	6.7	6.7	1.8	1.8
Sometimes unemployed	19.9	20.5	15.2	21.1
Out of labor force	1.6	1.6	1.1	1.1
Number	312	312	171	171

The results which have been cited for the pre-migration period were found to be substantially unaltered when separate tabulations were made excluding the migrants with a farm background.

The peak unemployment rate for migrants occurred after rather than before the migration. While the frequency of unemployment among migrants rose sharply in the post-migration period, the unemployment rate for the non-migrant controls showed only a moderate increase at either end of the movement. This is demonstrated by the data in Tables 1–4 to be true whether or not the totally unemployed are included in the pre-migration unemployed group. For example, while the unemployment rate for migrants to Flint increased from 24 to 49 per cent, that for the non-migrant source controls increased from 18 to 24 per cent. While the percentage of increase in unemployment rate was no more than 36 per cent for any non-migrant group, it was no less than 82 per cent for any migrant group. In every case the increase in unemployment rates for the migrant group was significantly

greater[4] than the corresponding increase for its non-migrant control—whether source or destination.

TABLE 3

PERCENTAGE DISTRIBUTION OF EMPLOYMENT STATUS OF MIGRANTS AND NON-MIGRANTS MATCHED AT DESTINATION FOR POST-MIGRATION PERIOD FLINT AND GRAND RAPIDS

Employment Status in Post-migration Period	Migrant Status and 1935 Residence			
	Flint		Grand Rapids	
	Migrant	Non-migrant	Migrant	Non-migrant
Total percentage	100.0	100.0	100.0	100.0
In labor force:	96.8	97.8	94.1	98.2
Employed	48.4	71.2	57.9	68.4
Unemployed	48.4	26.6	36.2	29.8
Out of labor force	3.2	2.2	5.9	1.8
Number	312	312	171	171

TABLE 4

PERCENTAGE DISTRIBUTION OF EMPLOYMENT STATUS OF MIGRANTS AND NON-MIGRANTS MATCHED AT SOURCE FOR POST-MIGRATION PERIOD FLINT AND GRAND RAPIDS

Employment Status in Post-migration Period	Migrant Status and 1935 Residence			
	Flint		Grand Rapids	
	Migrant	Non-migrant	Migrant	Non-migrant
Total percentage	100.0	100.0	100.0	100.0
In labor force:	97.6	98.6	96.0	97.4
Employed	48.6	74.3	58.4	82.6
Unemployed	49.0	24.3	37.6	14.8
Out of labor force	2.4	1.4	4.0	2.6
Number	296	296	149	149

The relatively high post-migration unemployment rate was not restricted to a particular occupational group, although the migrants with a farm background had the highest unemployment rates. Tables 5 and 6 show the unemployment status of each migrant and non-migrant control group separately for 3 major occupational groupings:

[4] The differences were significant at the .01 level in every case but one. The difference for the Grand Rapids migrant and non-migrant destination groups was significant at the .05 level. The differences were significant at these levels whether or not the "totally" unemployed were considered in the pre-migration period.

TABLE 5

PERCENTAGE DISTRIBUTION OF POST-MIGRATION EMPLOYMENT STATUS BY OCCUPATIONAL CLASS, MIGRANTS AND NON-MIGRANTS MATCHED AT SOURCE, FLINT AND GRAND RAPIDS, MICHIGAN

Occupational Class before Migration	Flint						Grand Rapids					
	Migrants			Non-migrants			Migrants			Non-migrants		
	Total[1]	Employed	Unemployed	Total[1]	Employed	Unemployed	Total[1]	Employed	Unemployed	Total[1]	Employed	Unemployed
Total	100.0	50.0	50.0	100.0	75.3	24.7	100.0	64.7	35.3	100.0	85.4	14.6
White collar[2]	100.0	77.0	23.0	100.0	85.3	14.7	100.0	75.4	24.6	100.0	94.8	5.2
Blue collar[3]	100.0	49.0	51.0	100.0	69.5	30.5	100.0	61.4	38.6	100.0	71.1	28.9
Farm workers[4]	100.0	22.5	77.5	100.0	86.4	13.6	100.0	46.4	53.6	100.0	87.1	12.9
Others[5]	100.0	—	—	100.0	—	—	100.0	—	—	100.0	—	—

[1] Exclusive of individuals out of the labor force during entire post-migration period.
[2] Includes: professional workers; proprietors, managers, officials; and clerical workers.
[3] Includes: skilled, semi-skilled, and unskilled workers, and servants.
[4] Includes: farmers and farm laborers.
[5] Includes all persons unemployed or out of labor force during entire pre-migration period. Numbers too small for computation of percentages.

TABLE 6

PERCENTAGE DISTRIBUTION OF POST-MIGRATION EMPLOYMENT STATUS BY OCCUPATIONAL CLASS, MIGRANTS AND NON-MIGRANTS MATCHED AT DESTINATION, FLINT AND GRAND RAPIDS, MICHIGAN

Occupational Class	Flint						Grand Rapids					
	Migrants			Non-migrants			Migrants			Non-migrants		
	Total[1]	Employed	Unemployed	Total[1]	Employed	Unemployed	Total[1]	Employed	Unemployed	Total[1]	Employed	Unemployed
Total	100.0	50.0	50.0	100.0	72.8	27.2	100.0	62.4	37.6	100.0	70.6	29.4
White collar[2]	100.0	77.0	23.0	100.0	94.5	5.5	100.0	75.0	25.0	100.0	80.9	19.1
Blue collar[3]	100.0	45.4	54.6	100.0	68.8	31.2	100.0	58.5	41.5	100.0	61.8	38.2
Farm workers[4]	100.0	25.6	74.4	100.0	74.4	25.6	100.0	41.7	58.3	100.0	65.0	35.0
Others[5]	100.0	—	—	100.0	—	—	100.0	—	—	100.0	—	—

[1] Exclusive of individuals not in the labor force.
[2] Includes: professional workers, proprietors, managers, officials, and clerical workers.
[3] Includes: skilled, semi-skilled, and unskilled workers, and servants.
[4] Includes: farmers and farm laborers.
[5] Includes all persons unemployed or out of labor force during entire pre-migration period. Numbers too small for computation of percentages.

TABLE 7

PERCENTAGE DISTRIBUTION OF POST-MIGRATION EMPLOYMENT STATUS, BY EDUCATION, MIGRANTS AND NON-MIGRANTS MATCHED AT SOURCE, FLINT AND GRAND RAPIDS, MICHIGAN

Highest Educational Level Reached	Flint						Grand Rapids					
	Migrants			Non-migrants			Migrants			Non-migrants		
	Total[1]	Em-ployed	Unem-ployed	Total[1]	Em-ployed	Unem-ployed	Total[1]	Em-ployed	Unem-ployed	Total[1]	Em-ployed	Unem-ployed
Total	100.0	49.5	50.5	100.0	75.2	24.8	100.0	64.2	35.8	100.0	85.4	14.6
Grade school	100.0	33.3	66.7	100.0	73.5	26.5	100.0	53.3	46.7	100.0	84.4	15.6
High school	100.0	67.3	32.7	100.0	76.0	24.0	100.0	73.7	26.3	100.0	82.0	18.0
College	100.0	79.2	20.8	100.0	83.5	16.5	100.0	89.5	10.5	100.0	95.2	4.8

[1] Exclusive of individuals out of labor force during entire post-migration period.

TABLE 8

PERCENTAGE DISTRIBUTION OF POST-MIGRATION EMPLOYMENT STATUS, BY EDUCATION, MIGRANTS AND NON-MIGRANTS MATCHED AT DESTINATION, FLINT AND GRAND RAPIDS, MICHIGAN

Highest Educational Level Reached	Flint						Grand Rapids					
	Migrants			Non-migrants			Migrants			Non-migrants		
	Total[1]	Em-ployed	Unem-ployed	Total[1]	Em-ployed	Unem-ployed	Total[1]	Em-ployed	Unem-ployed	Total[1]	Em-ployed	Unem-ployed
Total	100.0	50.0	50.0	100.0	72.8	27.2	100.0	64.2	35.8	100.0	72.9	27.1
Grade school	100.0	35.3	64.7	100.0	60.7	39.3	100.0	53.2	46.8	100.0	65.4	34.6
High school	100.0	65.8	34.2	100.0	75.7	24.3	100.0	73.7	26.3	100.0	78.1	21.9
College	100.0	69.2	30.8	100.0	100.0	0.0	100.0	89.5	10.5	100.0	90.5	9.5

[1] Exclusive of individuals out of labor force during entire post-migration period.

UNIVERSITY OF MICHIGAN LIBRARIES

white collar workers (professional workers, proprietors and managers, officials, and clerical workers), blue collar workers (skilled, semi-skilled, and unskilled workers, and servants), farm workers (farmers and farm laborers). In each of these occupational categories the post-migration unemployment rate of the migrants to Flint or Grand Rapids was higher than that for comparable non-migrant control groups at the source or destination. It is clear from these data that the farm migrants had a much higher rate of unemployment than any other group whether migrant or non-migrant.

Similarly, the relatively high unemployment rate of the migrants is found at each of three educational levels. In Tables 7 and 8 the employment status of each migrant and non-migrant group is shown separately for those with grade school, high school, and college education, respectively. The only comparison in which the migrants do not have the higher rate of unemployment is that between migrants to Grand Rapids and non-migrants to Grand Rapids for those with some college education. The numbers involved in this comparison are very small.

The sizes of the sub-samples of the occupational groups and educational groups on which the percentages for Tables 5–8 are based are found in Table 9.

TABLE 9

NUMBER IN EACH SAMPLE, BY OCCUPATIONAL CLASS AND EDUCATIONAL LEVEL, FLINT AND GRAND RAPIDS

Occupational Class and Educational Level	Flint		Grand Rapids	
	Sample Matched at Source	Sample Matched at Destination	Sample Matched at Source	Sample Matched at Destination
Occupational class: Total	296	312	149	171
White collar	88	94	60	69
Blue collar	106	113	48	56
Farm	83	79	39	41
Others	19	26	2	5
Educational level: Total	296	312	149	171
Grade school	166	168	86	98
High school	106	117	43	49
College	24	27	20	24

It is recognized that for these various post-migration comparisons it would have been desirable that the matching include control on pre-migration employment status. However, it can be shown that this

matching defect does not seriously affect the findings. In the first place, the difference in *increase* in unemployment rates (not just the absolute post-migration differences in unemployment) as between migrant and non-migrant is significantly greater for migrants in every comparison. Secondly, separate post-migration unemployment rates were computed for those migrants unemployed and those never unemployed in the pre-migration period. In every case both the previously employed and previously unemployed migrant groups had higher post-migration unemployment rates than the non-migrant controls. The difference in unemployment rates as between migrants with a pre-migration history of unemployment and those with none was relatively small.

In this connection it is interesting that the correlation between pre-migration and post-migration unemployment was considerably greater for the non-migrants than for their migrant counterparts. Among the migrants the rate of post-migration unemployment was high both for those who had and those who did not have pre-migration unemployment. On the other hand among the non-migrants the rate of post-migration unemployment was high for those with pre-migration unemployment but very low for those without it. Apparently, for migrants the fact of migration was more important than previous employment status in determining employment status in their new home.

Consideration of the length of unemployment for the unemployed in each group of migrants and non-migrants indicates that the higher unemployment rate of the migrants is not a consequence of short unemployment periods distributed among many persons. Among those unemployed after migration, the average length of unemployment was greater for migrants than for non-migrants in every comparison but one, and in that case the average length of unemployment was the same for each group. This is in part an answer to the idea that the migrant unemployment was of a short-run "frictional" character.

Whatever the effects on the sending and receiving communities, migration appears to have offered no effective solution to unemployment for the migrants, themselves. Quite the contrary, the evidence seems to support the thesis that migration resulted in a deterioration of their employment status. At least for the short-run periods considered in this study, if there is any line of causation between migration and unemployment it appears to run from migration to unemployment rather than the reverse.

Evidence for the thesis that unemployment "causes" migration is sometimes presented in the form of data showing that over a given period of time the unemployment rate was highest for those who

migrated during the period. Thus, it may be said that 60 per cent of the source-matched migrants were unemployed sometime between 1930 and 1935, as compared with only 27 per cent of matched-source non-migrants. Such evidence is really not relevant to the hypothesis, since it fails to indicate whether the unemployment preceded or followed the migration. Although it is true that 60 per cent of the Flint migrants[5] were unemployed at some time during the 5 year period, only 23 per cent of the migrants as compared with 18 per cent of the source non-migrants were unemployed prior to migration. The large differential in unemployment rates is *after* migration, not *before*.

The results of the present study are consistent with the hypothesis that in a depression migrants tend to be at a disadvantage in the new labor market to which they move. This may result from the fact that the migrants lack the specific skills demanded in the local labor market. It may be that, although they have the needed skills, the migrants are at a disadvantage in their poorer knowledge of how to get and hold a job in the local situation, or it may be that in a depression employers are under pressure to favor the local labor supply.

The methodological aspects of this study are equally important with the substantive findings. At least in this instance, the use of matched group control techniques proved to be feasible in a migration study involving a relatively small group. The technique furnishes factor-control frequently lacking in this kind of study. Further experimentation with the technique appears to be warranted.

[5] Sixty per cent of the source-matched migrants and 61 per cent of the destination-matched migrants.

MINIMUM X^2 AND MAXIMUM LIKELIHOOD SOLUTION IN TERMS OF A LINEAR TRANSFORM, WITH PARTICULAR REFERENCE TO BIO-ASSAY

JOSEPH BERKSON, M.D.
Division of Biometry and Medical Statistics, Mayo Clinic, Rochester, Minnesota

An old and commonly used device for fitting a curve $y=F(x, \alpha, \beta)$ is to find a function of y which is linearly related to x, or a function of x which is linearly related to y, or functions of y and x which are linearly related to each other. The linearly related functions are plotted against one another, and a line is fitted to the points, usually by eye, sometimes by "least squares" in terms of the transforms. Texts dealing with empirical curve-fitting characteristically use this scheme [6]. Before Bliss and Fisher [2], no attempt was made to adjust the transforms systematically in order to achieve a fit that fulfilled defined criteria in terms of the original measures y and x, nor was it known that such adjustments were possible. These authors presented for the bio-assay experiment a method in terms of probits, which as shown by Garwood, [5] accomplished a maximum likelihood estimate of the integrated normal curve, when the observations were distributed binomially. Following the procedures used by Bliss and Fisher for the integrated normal curve, I [1] formulated the similar adjustments applicable to "logits" for a maximum likelihood solution of the logistic function, but did not advocate using this solution. Finney [4] has recently presented the adjustments for a maximum likelihood solution of several functions of interest in bio-assay. So far as I know the similar method for a solution fulfilling the criteria of minimum X^2 rather than maximum likelihood has not been presented. In the following paper this is given for several functions in common statistical use, and at the same time a résumé is given of the maximum likelihood adjustments for the same functions.

WE ARE DEALING with a measure Q $(P=1-Q)$ which is a function of x and two parameters α, β, the estimates of these parameters a, b, and Y, a linear transform of Q.

$$Q_i = F(x_i, \alpha, \beta) = F(Y_i) \tag{1}$$

$$Y_i = \alpha + \beta x_i \tag{2}$$

$$\hat{q}_i = F(x_i, a, b) = F(\mathcal{Y}_i) \tag{3}$$

$$\mathcal{Y}_i = a + bx_i. \tag{4}$$

Let

$$z = \frac{\partial \hat{q}}{\partial \hat{y}}\ ; \quad \text{then} \quad \frac{\partial \hat{q}}{\partial a} = \hat{z}, \quad \frac{\partial \hat{q}}{\partial b} = \hat{z}x.$$

If the observation q_i is distributed normally and independently with σ equal at all values of x, then the estimates a and b fulfilling the criteria of either maximum likelihood or minimum X^2 are identical, and are given by the normal equations

$$\sum n_i \hat{z}_i (q_i - \hat{q}_i) = \sum w_i'(q_i - \hat{q}_i) = 0 \tag{5*}$$

$$\sum n_i \hat{z}_i x_i (q_i - \hat{q}_i) = \sum w_i' x_i (q_i - \hat{q}_i) = 0. \tag{6}$$

If q_i is distributed binomially with $\sigma_i^2 = (P_i Q_i)/n_i$, the normal equations for maximum likelihood are given by

$$\sum \frac{n_i}{\hat{p}_i \hat{q}_i} (q_i - \hat{q}_i) \frac{\partial \hat{q}_i}{\partial a} = \sum \frac{n_i}{\hat{p}_i \hat{q}_i} \hat{z}_i (q_i - \hat{q}_i)$$

$$= \sum w_i'' (q_i - \hat{q}_i) = 0 \tag{7}$$

$$\sum \frac{n_i}{\hat{p}_i \hat{q}_i} (q_i - \hat{q}_i) \frac{\partial \hat{q}_i}{\partial b} = \sum \frac{n_i}{\hat{p}_i \hat{q}_i} \hat{z}_i x_i (q_i - \hat{q}_i)$$

$$= \sum w_i'' x_i (q_i - \hat{q}_i) = 0. \tag{8}$$

For X^2 we have

$$X^2 = \sum \frac{n_i}{\hat{p}_i \hat{q}_i} (q_i - \hat{q}_i)^2$$

and to minimize we set the first derivatives with respect to a and b equal to zero, yielding

$$\frac{\partial X^2}{\partial a} = -2 \sum \frac{n_i}{\hat{p}_i \hat{q}_i} (q_i - \hat{q}_i) \frac{\partial \hat{q}_i}{\partial a}$$

$$- \sum \frac{n_i}{(\hat{p}_i \hat{q}_i)^2} (1 - 2\hat{q}_i)(q_i - \hat{q}_i)^2 \frac{\partial \hat{q}_i}{\partial a} = 0 \tag{9}$$

$$= \sum \frac{n_i}{(\hat{p}_i \hat{q}_i)^2} \hat{z}_i (\hat{p}_i q_i + p_i \hat{q}_i)(q_i - \hat{q}_i)$$

$$= \sum w_i''' (q_i - \hat{q}_i) = 0 \tag{10}$$

* The primes on the w's are to distinguish between the different weights; they do not represent successive differential coefficients.

and similarly for b

$$\sum \frac{n_i}{(\hat{p}_i\hat{q}_i)^2}\hat{z}_i(\hat{p}_iq_i + p_i\hat{q}_i)x_i(q_i - \hat{q}_i) = \sum w_i'''x_i(q_i - \hat{q}_i) = 0. \quad (11)$$

The condition for the minimum X^2 (9) differs from the condition (7) of maximum likelihood only in the second term of (9). This term will vanish as (a) $\hat{q}_i \rightleftarrows 0.5$, (b) $q_i \rightleftarrows \hat{q}_i$ so that for experiments such as bio-assay where observations are in the neighborhood of the L.D. 50, or where the observations are close to the curve the two solutions may be expected not to differ very much.[1]

It is seen that all the normal equations are of the form

$$\sum w(q - \hat{q}) = 0$$

or

$$\sum wq = \sum w\hat{q}.$$

It is interesting to note therefore that both minimum X^2 and maximum likelihood, for normal or binomial variation, impose the simple requirement that the weighted average of the estimates be equal to the weighted average of the observations, the only difference among them being the value of the weights w.

For the situations usually considered in bio-assay work, the normal equations cannot be solved directly because the functions usually employed are not linear in the parameters, and where q is assumed to vary binomially there is the additional reason that the weights are a function of the $\hat{q}$'s, the values to be estimated. The equations may be solved by a procedure of successive approximations by making the following substitutions. Let $\hat{q}_{i,1}$ be a preliminary estimate of $\hat{q}_i$, $\hat{y}_{i,1}$ the corresponding value of $\hat{y}_i$ and $\hat{z}_{i,1}$ the corresponding value of $\hat{z}_i$, and for brevity write $\hat{q}_1$ for $\hat{q}_{i,1}$, $\hat{y}_1$ for $\hat{y}_{i,1}$ and $\hat{z}_1$ for $\hat{z}_{i,1}$.

$$q_i - \hat{q}_i \equiv (q_i - \hat{q}_1) + (\hat{q}_1 - \hat{q}_i).$$

$(\hat{q}_1 - \hat{q}_i)$ is replaced by the first term in a Taylor's expansion

[1] It is sometimes stated that it has been proved that the minimum X^2 and maximum likelihood solutions converge to the same solution as $n \rightarrow \infty$. From the foregoing it seems to me that this is not necessarily so. It will be so if $q \rightleftarrows \hat{q}$ as $n \rightarrow \infty$, but this is not always the case. When the function we are fitting is the "true" curve, theoretically $(q - \hat{q}) \rightleftarrows 0$ with probability approaching 1 as $n \rightarrow \infty$, but not if we are fitting a function that is different from the "true" function, as is usually the case in statistical curve fitting. The same is true if we are fitting a line by least squares minimizing the squared residual of y, to observations of a random sample from a bivariate normal distribution, when there is an error in the observation of x.

TABLE 1
WORKING VALUES AND WEIGHTS

Function	Transform, $\hat{y}$	$\hat{z}=\frac{\partial \hat{q}}{\partial \hat{y}}$	Working value	Weight, w		
				Normal	Binomial	
				Max. lik.; min. X^2	Max. Lik.	Min. X^2
Integrated normal curve $\hat{q}=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{a+bx} e^{-t^2/2}$	Normal deviate, $\hat{t}$ or Probit $=\hat{t}+5$	$\hat{z}$, ordinate of the normal curve at $\hat{t}$ $\frac{1}{\sqrt{2\pi}}e^{-t^2/2}$	$\hat{t}_1+\frac{q-\hat{q}_1}{\hat{z}_1}$	$n\hat{z}_1^2$	$\frac{n\hat{z}_1^2}{\hat{p}_1\hat{q}_1}$	$\frac{n\hat{z}_1^2}{(\hat{p}_1\hat{q}_1)^2}\left(\hat{p}_1q+p\hat{q}_1\right)$
Logistic $\hat{q}=\frac{1}{1+e^{-(a+bx)}}$	Logit* $\hat{l}=\ln\frac{\hat{q}}{\hat{p}}$	$\hat{p}\hat{q}$	$\hat{l}_1+\frac{q-\hat{q}_1}{\hat{p}_1\hat{q}_1}$	$n(\hat{p}_1\hat{q}_1)^2$	$n\hat{p}_1\hat{q}_1$	$n(\hat{p}_1q+p\hat{q}_1)$
Sine $\hat{q}=\sin^2(a+bx)$	Unchristened $\hat{k}=\sin^{-1}\sqrt{\hat{q}}$	$2\sqrt{\hat{p}\hat{q}}$	$\hat{k}_1+\frac{q-\hat{q}_1}{2\sqrt{\hat{p}_1\hat{q}_1}}$	$n\hat{p}_1\hat{q}_1$	n	$\frac{n(\hat{p}_1q+p\hat{q}_1)}{\hat{p}_1\hat{q}_1}$
Exponential $\hat{q}=e^{a+bx}$	Natural logarithm $\hat{h}=\ln\hat{q}$	$\hat{q}$	$\hat{h}_1+\frac{q-\hat{q}_1}{\hat{q}_1}$	$n\hat{q}_1^2$	$\frac{n\hat{q}_1}{\hat{p}_1}$	$\frac{n(\hat{p}_1q+p\hat{q}_1)}{\hat{p}_1^2}$

* The sign of the "logit" as given here is the negative of that as defined previously [1].

$$(\hat{q}_1 - \hat{q}_i) \doteq \hat{z}_1(\mathcal{Y}_1 - \mathcal{Y}_i)$$

$$(q_i - \hat{q}_i) \doteq (q_i - \hat{q}_1) + \hat{z}_1(\mathcal{Y}_1 - \mathcal{Y}_i)$$

$$\doteq \hat{z}_1\left[\left(\frac{q_i - \hat{q}_1}{\hat{z}_1} + \mathcal{Y}_1\right) - \hat{y}_i\right]$$

$$\doteq \hat{z}_1(\mathcal{Y}_i' - \mathcal{Y}_i). \tag{12}$$

The value of $\mathcal{Y}_i' = (q_i - \hat{q}_1)/(\hat{z}_1) + \mathcal{Y}_1$ is called the "working value." If we replace $(q_i - \hat{q}_i)$ as in (12) in any of the normal equations previously given, and, where in any of those equations $\hat{z}_i$ appears, substitute the preliminary approximation $\hat{z}_1$, since $\mathcal{Y}$ is linear in a and b, the equations may be solved by the usual procedures as used for fitting a

TABLE 2

FIT OF LOGISTIC FUNCTION*

Iteration	Minimum X^2				Maximum likelihood			
	a	b	X^2	$L+30$†	a	b	X^2	$L+30$†
Original‡	−2.972	2.515	3.9480	0.7988	−2.972	2.515	3.9480	0.7988
1	−3.156	2.473	2.8404	1.0278	−3.290	2.581	2.9943	1.0644
2	−3.140	2.467	2.8390	1.0242	−3.324	2.602	3.0658	1.0657
3	−3.143	2.469	2.8389	1.0255	−3.325	2.602	3.0666	1.0657
4	−3.143	2.469	2.8389	1.0252	−3.325	2.602	3.0666	1.0657
Approximate minimum X^2, noniterative method								
	−3.126	2.444	2.8501	1.0074				

* Example from Finney [3]. To simplify the calculations x, the log dose, as given by Finney was adjusted to make x = zero for the lowest dose, and since the doses proceeded in multiples of 2, it was divided by log 2. The iterations were continued until the absolute difference between successive values of both a and b was less than 0.0005.

† $L = \Sigma r \log \hat{p}_i + \Sigma(n_i - r_i) \log \hat{q}_i$, where n_i is the number exposed at dose x_i, p_i is the fraction affected at x_i, $r_i = p_i n_i$ and $\hat{p}_i$ is the solution value of $p_i(\hat{q}_i = 1 - \hat{p}_i)$.

‡ Solution using n alone as weight.

straight line by least squares, using as weight $w_i = w_i'\hat{z}_1$, $w_i = w_i''\hat{z}_1$ or $w_i = w_i'''\hat{z}_1$ as the case may be. With a solution obtained in this manner, the procedure outlined is repeated, using for the evaluation of the new weights and working values, the values of $\hat{q}$ obtained in the present solution, and by repeating this, the respective minimum X^2 or maximum likelihood solution is approached as a limit.

In Table 1 are given for some functions commonly used in statistical practice the working values and weights in the case of normal and bi-

nomial variation, for the minimum X^2 and maximum likelihood solutions. In Table 2 are given the results of fitting a logistic function to some data used by Finney [3] by the methods outlined for (1) minimum X^2; (2) maximum likelihood; (3) noniterative approximate minimum X^2 solution previously suggested by me [1].

REFERENCES

[1] Berkson, J., "Application of the Logistic Function to Bio-Assay" (1944). *Journal of the American Statistical Association*, 39, 357.

[2] Bliss, C. I., with an Appendix by Fisher, R. A., "The Calculation of the Dosage Mortality Curve" (1935). *Annals of Applied Biology*, 22, 134.

[3] Finney, D. J., "Probit Analysis: A Statistical Treatment of the Sigmoid Response Curve." Cambridge University Press, 1947, p. 201.

[4] Finney, D. J., "The Principles of Biological Assay" (1947). *Journal of the Royal Statistical Society, (Suppl.)* 9, 46.

[5] Garwood, F., "The Application of Maximum Likelihood to Dosage Mortality Curves" (1940). Biometrika, 32, 46.

[6] Lipka, Joseph, "Graphical and Mechanical Computation." Ed. 1, New York, John Wiley & Sons, Inc., 1918, 264 pp.

SOME INADEQUACIES OF THE FEDERAL CENSUSES OF AGRICULTURE*

RAYMOND J. JESSEN
Bureau of Agricultural Economics, and Statistical Laboratory, Iowa State College

INTRODUCTION

A NUMBER OF criticisms of the federal censuses of agriculture have been made by users of census data. The primary purpose of this paper is to bring together for discussion some of these criticisms, and to propose modifications designed to make the censuses more adequate and useful (i) as sources of general statistics on agriculture and (ii) as sources of information for economic analyses. A secondary purpose is to consider some questions of survey methods for obtaining agricultural information—whether the survey be a census or a sample. The results of a recent statistical survey of the United States[1] are used as an aid in determining the general order of magnitude of some of the points under consideration.

The definition of a farm used by the Bureau of the Census appears to be inadequate for the purposes of agricultural economists on two counts: the exclusion of certain agricultural producing units because of (i) the *scale of operations* and (ii) the *type of enterprise*. In this connection it is estimated from the survey that the 1945 Census of Agriculture excluded, by the definition of a farm adopted, a group of about 5,200,000 units whose scale of agricultural operations was too small to qualify them as "census farms" but which contributed an estimated $445,000,000, or about 1.8%, to total agricultural production in 1946. This in itself may not be important, but where *kinds* of agricultural production are under examination it may be of practical significance.

In addition to the omissions from the census of units that do not meet the definition of a farm, there is some evidence that the census may have failed to enumerate some farms that do meet the definition. This is true particularly for comparatively small farming operations. The apparent deficiencies in the coverage of farms by the census may have arisen because of an incompleteness of coverage by the census inves-

* Journal paper No. J-1635 of the Iowa Agricultural Experiment Station, Ames, Iowa, Project 1005, in cooperation with the Bureau of Agricultural Economics, U. S. Department of Agriculture. The views and conclusions expressed in this paper are entirely those of the author and do not necessarily in any way reflect the views of the Bureau of Agricultural Economics.

[1] A national survey of the household refrigeration market conducted by the Statistical Laboratory in January, 1947. This survey was not designed for this purpose but it does have some information which may be of interest here.

tigator, or because of differences in the interpretation of the value of products of certain operations by the census investigator and by the investigators in the refrigeration survey.

In connection with the problem of choice of area smaller than the county for presentation of data, it is suggested that a "natural" area such as a valley or some such drainage area be adopted by the Bureau of the Census, rather than the present unit which is mainly political and indeed subject to change from time to time. This would provide statistics on a more useful and permanent basis almost throughout the U.S., and particularly in the west where political divisions are of especially small agricultural utility.

General discussion of the problem. The data from federal government censuses are used by a great many persons and for a great many purposes. For the present discussion criticism will be considered mainly from the standpoint of its use for research in agricultural economics. Workers in this field frequently make use of the total farm count, area under various crops, inventories of a number of farm quantities, characteristics of the farm dwelling, etc. Agricultural statisticians, such as those of agricultural Estimates, Bureau of Agricultural Economics, use the data as a "bench mark" on which to base a large body of current estimates on acreages under different crops, crop yields, livestock numbers and production, etc.

Three classes of users of census data in agricultural economics will be distinguished: (1) those who are interested in the number and characteristics of "farms" as firms, (2) those who are interested in the number and characteristics of "farmers" and their families and (3) those who are interested in statistical aggregates—the total production of rice, for example. (There are of course other classes such as those interested in land owners, for example, but only these three classes will be dealt with here.) The first group requires data to be given in the form of frequency tables in which the farm is the unit (examples: Number of farms by size in acres, and by value of output). The second group requires the data also in frequency tables but with the farmer (operator) as the unit (example: income of farm operators by age group). The third group is generally not concerned with the unit by which an item is collected or tabulated but merely the unit of measurement (bushel, head, etc.). It is possible, however, that there is a demand on the part of this group for such frequency distributions as acreage of harvested corn by size of field, yield of apples by age of tree, etc.

The conflict between *farm* and *farmer* as a unit of enumeration and

tabulation (class 1 versus class 2) is of real importance because many farms are "operated" by more than one farmer—for example, partnerships.

In attempting to satisfy the demands of all classes of users, it was of course necessary for the Bureau of the Census to make some compromises which, in part at least, gave rise to the inadequacies to be discussed here. Consideration will be given to those inadequacies arising from:

(1) definition of farm
(2) enumeration
(3) definition of farmer
(4) definition of agriculture
(5) unit for presentation of data for small areas.

Inadequate definition of farm. In order to discuss properly the implications of the "farm" as regarded by the Bureau of the Census, it may be appropriate to quote in full the definition given the enumerators in the 1945 census of agriculture. This definition follows.

> A farm, for Census purposes, is all the land on which some agricultural operations are performed by one person, either by his own labor alone or with the assistance of members of his household, or hired employees. The land operated by a partnership is likewise considered a farm. A "farm" may consist of a single tract of land, or a number of separate tracts, and the several tracts may be held under different tenures, as when one tract is owned by the farmer and another tract is rented by him. When a landowner has one or more tenants, renters, croppers, or managers, the land operated by each is considered a farm. Thus, on a plantation the land operated by each cropper, renter, or tenant should be reported as a separate farm, and the land operated by the owner or manager by means of wage hands should likewise be reported as a separate farm.
>
> Include dry-lot or barn dairies, nurseries, greenhouses, hatcheries, fur farms, mushroom cellars, apiaries, cranberry bogs, etc.
>
> Do not include "fish farms", "fish hatcheries", "oyster farms", and "frog farms". *Do not report as a farm any tract of land less than 3 acres, unless its agricultural products in 1944 were valued at $250 or more.*
>
> Farming, or agricultural operations, consists of the production of crops or plants, vines, and trees (excluding forestry operations) or of the keeping, grazing, or feeding of livestock for animal products (including forestry operations) or of the keeping, grazing, or feeding of livestock for animal products (including serums), animal increase, or value increase. Livestock, as here used, includes poultry of all kinds, rabbits, bees, and fur-bearing animals in captivity—in addition to mules, asses, burrows, horses, cattle, sheep, goats, and hogs. Frequently certain operations are not generally recognized as farming. This is especially true where no crops are grown or where the establishments are not commonly considered as farms.

Following is a partial list of types of specialized agriculture and of operations not generally recognized as farms or farming, for which returns on the Farm and Ranch Schedule are required, provided the area is 3 acres or more or, if less than 3 acres, the value of the products in 1944 was $250 or more: Apiaries (bee farms), Community or cooperative gardens, Country estates and country homes (if there is production of vegetables, eggs, milk, or other agricultural products either for home use or for sale), Cranberry bogs, Dry-lot or barn dairies, Feed lots, Fur farms (fox, mink, skunk, etc., in captivity), Garbage-feeding hog yards, Greenhouses, Hatcheries (baby chicks, poults, etc.), Institutional farms (connected with schools, prisons, hospitals, etc.) Mushroom cellars, Nurseries (except for reforestation projects, or in connection with parks), Part-time farms (agricultural operations incidental to other occupation), Grazing or pasturing of livestock, Harvesting of grass seed, Keeping of chickens and the production of broilers (including battery-laying and battery-broiler plants), Production of medicinal or drug plants and herbs, Production of flowers and bulbs for sale, Production of vegetables under glass, Production of vegetable and flower seeds, plants, bulbs, tubers, etc., Production in captivity of pheasants, quail, etc., Production of mint, spices, or other special crops, Raising of Shetland or other ponies, Rabbit raising, Squab raising.

If any specialized or unusual types of agriculture such as those mentioned above are reported, list type under Supplemental Information on page 12.

Although columns are not provided on the schedule for obtaining reports for all the above-mentioned specialized operations in detail, be sure to report on all items that are applicable, making use of inquiries, for "other crops". Note that value of land and buildings and value of sale of products should be reported in all cases.

Include in one report all such land which the operator uses for agricultural purposes, as previously defined, also all outlying or separate field, meadows, pastures, woodland, and waste land. A farm may consist of two or more separate tracts not necessarily adjacent. Do not include public or open range neither owned nor leased by the operator. If the operator cuts hay from land that he does not own and for which he pays no rent, include such acreage under Wild Hay Cut and explain under supplemental information. Large areas of land or other non-agricultural land held as a separate business and not used for pasture or grazing should not be included.

The following types of establishments and operations do not require returns on the Farm and Rranch schedules unless there are also agricultural operations: Canneries, Cheese factories, Creameries, Deer parks, Fish, frog, alligator, or snake "farms", Fish hatcheries, Game preserves, Kennels, Livestock dealers (except feed lots or other farming operations), Ostrich "farms", Oyster "farms" Parks, Riding academies with no farming operations, Shipping pens, Turpentine "farms" or turpentine "orchards", Distilleries, gins, dryers, mills, refineries, or packing plants, Establishments of 3 acres or more, even though locally known as "farms" on which there are no agricultural operations. Idle or abandoned farms which were not operated in 1944 and will not be operated in 1945, Cutting or gathering of forest products with no farming operations, Landscaping, or maintaining grounds,

and growing of flowers, shrubs, and ornamental etc., except where the land is maintained primarily for their production. Production of maple sirup or sugar with no farming operations, raising canaries, guinea pigs, white rats, or white mice, stock yards and auction yards or barns, trapping of wild animals.

It will be observed that the definition given above excludes certain agricultural producing units because of (i) scale of operations (as measured in acres and in dollars) and other elements are excluded because of (ii) *type* of *enterprise*. Later (Section 6) the effects of type of enterprise will be dealt with when the question of the scope of agriculture is being considered.

An estimate of the *amount* of agricultural production excluded by this definition because of insufficient scale is provided by a recent national sample survey.[3] According to this survey the total agricultural production (stated by producers)[4] was about $25 billion in 1946 (table 1). Of this quantity, $445 million, or 1.8% took place on "subcensus" farms (those too small to qualify as census farms). In general, units large enough to qualify as census farms and whose "operators" reside in the open country, produce crops and animals in about the same dollar volume, but all smaller units whether operated in the cities, towns or villages produce crops valued at about 5 times that of animals (Table 2).

[3] The National Refrigeration Survey dealt with a cross-sectional sample of households both "farm" and "non-farm". A special attempt was made to get households to report agricultural production however small, by avoiding questions referred to *farm* or *ranch* when the respondent said he had no *farm*. In these cases the interviewer was instructed to tell his respondent that even though he did not have a *farm* he might nevertheless be engaged in the production of some crops and animals on his *place* and proceeded with the interview asking for details on such activities. In the office these data were used to classify those units which qualified, as "census farms," the smaller producers as "subcensus farms". A weakness in the survey was the inadequate means for detecting multi-operator farms and therefore their inclusion in the sample.

[4] Includes production used at home. It is realized, of course, that this information reported by the respondent, is subject to error due to vagueness in concept, memory, etc.

The question, put to all households, whether they appeared to engaged in agricultural operations or not, were:

"Does head of household operate a farm?"

(if yes) "How much would you say the garden and field crops sold last year were worth at local market price? (Include home consumed as well as sold.)
How much would you say the animals sold or the products obtained from them and sold last year (such as milk, eggs, furs, honey) were worth at local market price? (Include home consumed as well as sold.)

(if no) "Even though you don't have a farm or regard yourself as a farmer, do you have a garden, chickens or some other small enterprise for production of agriculture products?

(If, No) no further agricultural questions.

(If, Yes) "How much were garden vegetables, fruits, berries etc. worth at the local market price? (Include home consumed as well as sold.)
"How much would you say that animals or animal products were worth at local market price? (Include home consumed as well as sold.)
(Also questions on acreage, etc.)

TABLE 1

ESTIMATED NUMBER AND VALUE OF AGRICULTURAL PRODUCTION IN 1946, OF FARMS AND "SUBFARMS", U. S. BY ZONE

(1)	(2)	(3)	(4)
Zone[a]	Estimated "census" & "subcensus" farms (that is all units reporting some agricultural production)[b]	Estimated "census" farms (those which qualify as farms according to census definition) and relative standard error[c]	Estimated "subcensus" farms (those excluded from the census because scale too small) and relative standard error (2)–(3)
Total			
Number	11,788,000	6,582,000 ± 5%	5,207,000 ± 12%
Production	$24,957,000,000	$24,512,000,000 ± 10%	$445,000,000 ± 12%
Open Country			
Number	5,946,000	5,200,000 ± 5%	746,000 ± 18%
Production	$19,784,000,000	$19,706,000,000 ± 11%	$ 77,000,000 ± 18%
Rural Place			
Number	2,857,000	890,000 ± 13%	1,966,000 ± 17%
Production	$ 2,972,000,000	$ 2,787,000,000 ± 28%	$185,000,000 ± 17%
Urban Place			
Number	2,985,000	491,000 ± 39%	2,494,000 ± 22%
Production	$ 2,202,000,000	$ 2,018,000,000 ± 46%	$184,000,000 ± 22%

[a] Zone refers to where the operator lives—not necessarily where farm is situated.

[b] Estimates from National Refrigeration Survey.

[c] Probability is about ⅔ that the difference between the sample estimate and the *true value* (the value that would result from a 100% sample) will be within plus or minus *one standard error*.

It is estimated that there are about 5,207,000 subcensus units producing some agricultural products (this number is not very meaningful because of the obvious indefiniteness of "*no* production"). Nearly half, or 2,494,000 of the operators of these subcensus farms reside in urban towns and cities; another 1,966,000 or 40% of the operators live in rural towns and villages and the remaining 746,000 or 14% are in the open country.[5] The place of residence of the operator may not be the same as the location of the unit. No detail was obtained in the survey to indicate the kinds of agricultural activity in which the operators of these units are engaged, but considering their location and scale of operations it would be reasonable to conjecture that they produce

[5] The partitioning of the U. S. into three "zones" follows a *practice* made possible by and established in connection with the Master Sample of Agriculture. By "urban" zone is meant the area covered by cities and towns having a population, in the 1940 Census of Population of 2500 or more, or otherwise designated by the Bureau of the Census as "urban". The "rural place" zone consists of all named places having at least 100 inhabitants, and other areas with a population density of at least 100 persons per square mile, which are not included in the urban group. These places, whether incorporated or not have had boundaries described around them as part of the Master Sample Project operations. The "open country" zone includes the remaining area of the U. S. outside of the "urban" and "rural place" zones.

mainly poultry and eggs, milk, vegetables, fruits, nuts, rabbits and the like. These are items that may be of importance in food and diet statistics.

Not only does the census definition of farm exclude a segment of agriculture on the basis of scale of operations, but it does so without consistency through time. The total number of farms enumerated by the census, other things being equal, depends on the agricultural price level. During years of high agricultural prices a greater number of units should qualify as farms than during years of low agricultural prices—even though physical production remains the same. This is due to the fact that one of the criteria chosen to measure scale is agricultural production in terms of dollar value ($250 per year). This sensitivity of the census farm count to the price situation could give rise to unfortunate complexity in the interpretation of a time series of farm counts. Specu-

TABLE 2

VALUE OF CROP PRODUCTION AND LIVESTOCK PRODUCTION 1946, OF "SUBCENSUS" AND "CENSUS" FARMS, U.S. BY ZONE[a]

(1)	(2)		(3)	(4)	(5)
Zone	Source		Total Census and Subcensus Farms	Census Farms and Relative Standard Error	Subcensus Farms and Relative Standard Error
Total, U. S.	Total	No.	11,788,000	6,581,000 ± 5.4	5,207,000 ± 12.3
		Value	24,957,000,000	24,512,000,000 ± 9.9	445,000,000 ± 12.1
	Crops	No.	11,225,000	6,217,000 ± 5.5	5,008,000 ± 11.8
		Value	14,267,000,000	13,894,000,000 ± 14.5	373,000,000 ± 12.4
	Animals	No.	6,891,000	5,553,000 ± 6.1	1,338,000 ± 24.4
		Value	10,690,000,000	10,618,000,000 ± 10.6	72,000,000 ± 24.3
Open Country	Total	No.	5,946,000	5,200,000 ± 5.3	746,000 ± 17.7
		Value	19,783,000,000	19,706,000,000 ± 10.8	77,000,000 ± 18.1
	Crops	No.	5,587,000	4,913,000 ± 5.8	674,000 ± 19.2
		Value	10,252,000,000	10,191,000,000 ± 16.1	61,000,000 ± 19.2
	Animals	No.	4,901,000	4,663,000 ± 6.2	238,000 ± 40.9
		Value	9,531,000,000	9,515,000,000 ± 11.3	16,000,000 ± 36.4
Rural Place	Total	No.	2,857,000	890,000 ± 12.7	1,967,000 ± 16.6
		Value	2,972,000,000	2,787,000,000 ± 27.6	185,000,000 ± 17.1
	Crops	No.	2,783,000	843,000 ± 13.1	1,940,000 ± 16.9
		Value	2,110,000,000	1,955,000,000 ± 35.0	155,000,000 ± 18.4
	Animals	No.	1,219,000	620,000 ± 14.6	599,000 ± 29.7
		Value	861,000,000	832,000,000 ± 32.5	29,000,000 ± 26.5
Urban Place	Total	No.	2,985,000	491,000 ± 39.2	2,494,000 ± 22.3
		Value	2,202,000,000	2,019,000,000 ± 46.1	183,000,000 ± 22.4
	Crops	No.	2,855,000	461,000 ± 38.4	2,394,000 ± 22.0
		Value	1,905,000,000	1,748,000,000 ± 54.5	158,000,000 ± 22.0
	Animals	No.	771,000	270,000 ± 53.9	501,000 ± 50.8
		Value	298,000,000	271,000,000 ± 64.7	27,000,000 ± 52.4

[a] Estimated from the National Refrigeration Survey.

TABLE 3

ESTIMATED NUMBER AND VALUE OF AGRICULTURAL PRODUCTION 1947, OF CENSUS AND SUBCENSUS FARMS, U.S. BY SIZE (IN ACRES) CLASS[a]

(Numbers in thousands, value in $ millions)

(1)	(2)		(3)	(4)	(5)	(6)	(7)	(8)	(9)
Value Class	Total		0–3	3–29	30–99	100–179	180–259	260–499	500 & over
1– 99	No.	3,299	3,046	151	75	21			6
	Value	137	127	7	3	*			*
100– 249	No.	2,654	2,160	342	108	38	3	3	
	Value	401	318	57	18	7	*	1	
250– 399	No.	690	303	257	112	15	3		
	Value	221	93	85	38	4	1		
400– 499	No.	437	164	117	111	33	3	9	
	Value	182	67	48	46	15	2	4	
500– 999	No.	1,014	194	330	308	137	39	6	
	Value	693	127	223	211	104	24	4	
1,000– 1,999	No.	912	15	237	414	145	54	33	14
	Value	1,253	20	309	565	213	81	45	20
2,000– 4,999	No.	1,354	14	86	388	464	190	143	69
	Value	4,248	33	254	1,153	1,457	586	511	254
5,000– 9,999	No.	738	—	30	74	193	138	198	105
	Value	5,253	—	219	488	1,319	1,008	1,439	780
10,000–19,999	No.	531	—	32	50	78	71	127	173
	Value	6,968	—	383	660	1,048	960	1,736	2,181
20,000 & over	No.	159	—	33	18	25	11	27	45
	Value	5,601	—	1,293	1,020	589	355	746	1,598
Total	No.	11,788	5,896	1,615	1,658	1,149	512	546	412
	Value	24,957	785	2,878	4,202	4,756	3,017	4,486	4,833

[a] Data from National Refrigeration Survey based on a sample therefore subject to sampling variation. See tables 1 and 5 for indications of the size of these sampling errors.

* Less than $500,000.

lating with data available from the survey (Table 3) it appears that if the agricultural price level is doubled (that is roughly what took place between the 1940 and 1945 census) about 1,500,000 farms should appear in the 1945 Census which would not have qualified for the 1940 census, assuming of course, the number and physical volume of production of all units producing agricultural products remains the same. Also if the price level should drop to one-half the 1945 level about 500,000 farms would be "lost". These speculations assume of course that enumerations each time are complete and that the definition of a farm

used in 1945 was strictly adhered to. These assumptions are not completely true, as will be pointed out in section 4. This sensitivity could be largely eliminated if the volume of production requirement in terms of dollar value were put at a value level each time of census such that a constant physical volume requirement would be maintained. Also if farm counts were published separately by zones, such as the three used here (open country, urban place and rural place), it might be possible to confine most of the changes due to price level fluctuations to the tables for towns and cities—the open country being somewhat freer from such effects.

Inadequate enumeration. Another part of agriculture that has been omitted by the Bureau of the Census, but included in its definition, is that which has been missed for one reason or another by the field worker, deleted in editing, or missed for other reasons. The differences shown in Table 4 between the survey estimates for 1947 and the data

TABLE 4

NUMBER OF FARMS BY CRITERION OF CENSUS DEFINITION SATISFIED, NATIONAL REFRIGERATION SURVEY (1947) AND 1945 CENSUS OF AGRICULTURE

Class	National Refrigeration Survey, Jan. 1, 1947 (estimate and standard error)	Federal Census of Agriculture April 1, 1945
Total "Census" Farms, U.S.	6,582,000± 5%	5,859,000
Qualifying on:		
Value alone	690,000±20%	99,000
Acreage alone	747,000±20%	552,000
Both acreage and value	5,145,000± 6%	5,208,000

of the 1945 Census of Agriculture are within the range given by twice the sampling error of the survey estimates, for both the total number of farms and value of agricultural production (when adjustment is made for the difference in price level between 1944 and 1946). However, if the number of farms is considered by classes according to the criterion (acreage, value or both) of the Census definition satisfied, it appears that about 590,000 of the total difference of 723,000 between the National Refrigeration Survey and the 1945 Census of Agriculture occurs in the class of farms qualifying on value of products alone (Table 4). The difference for this class is statistically significant.

There are a number of possible explanations of this difference. A reasonable interpretation is that farms of this type are often missed in the census. It appears that these farms are mostly small in scale and, consequently, might not have looked like farms to the Census investigator.

UNIVERSITY OF MICHIGAN LIBRARIES

As a rule they were situated in areas generally regarded as "non-farm" or had no resident operator and therefore might more easily have been passed over (Table 5). On the other hand, it might be that they were recognized by, or known to, the enumerator, but in his estimation were so costly to seek out and interview that he decided to omit them.

It is reasonable to believe that at least part of the discrepancy between the NR Survey and the census may be due to differences in the elicitation of value of products.[6] That is, a census investigator might well have visited a unit and found that the value of products (as obtained by him) was not sufficient to qualify the unit as a census farm,

TABLE 5

ESTIMATED NUMBER OF FARMS BY ZONE: 1940 CENSUS OF AGRICULTURE AND RESULTS OF NATIONAL REFRIGERATION SURVEY*

(1)	(2)	(3)
Zone	1940 Census of Agriculture.[a]	Survey estimate[b] and its standard error[c]
Total	6,096,799	6,582,000 ± 5.4%
Open country	5,532,374	5,200,000 ± 5.5%
Rural places	491,062	891,000 ± 13.6%
Urban places	73,363	491,000 ± 38.3%

* The two sets of estimates (1940 and 1947) are not directly comparable because of the fact that the estimates for the 1940 Census of Agriculture are based on location of the farm, while those from the National Refrigeration Survey are based on residence of the farm operator. These latter estimates (1947) tend to indicate more farms in the "urban places" and "rural places" zones. Thus, in the 1945 Census of Agriculture 340,000 farm operators reported themselves as not living on the farm they operated. Although place of residence was not tabulated for these operators, most of them undoubtedly resided in urban or rural places, rather than elsewhere in the open country zone.

[a] These estimates were made in connection with the Master Sample Project before the 1945 data were available. None have been made on the more recent data. Since they are estimates they are subject to error. Zone refers to location of farms and not residence of farm operator as in the Survey Estimates.

[b] Estimated from the National Refrigeration Survey sample. The sampling error of the estimate is given along side each figure. Zone refers to where the operator lives—not necessarily to where farm is situated, as in the estimates of 1940.

[c] Probability is about 2/3 that the difference between the sample estimate and the true value (the value that would result from a 100% sample) will be within *plus* or *minus one standard error*.

while for the same unit a survey investigator might have obtained a reported value of products that would satisfy the census criterion. In the absence of information as to how well value of products was elicited in the census and in the survey, and partly also because of the changes in price level, it is not possible to say which of the two figures for number of farms is more nearly correct. That is immaterial, however, for

[6] It should be noted that the procedure and questions for arriving at the total value of farm products were different in the National Refrigeration Survey than in the 1945 Census of Agriculture.

this analysis. The striking feature is that it is possible for two sets of presumably qualified investigators to take the same definition out in the field, and bring back significantly different results.

"Correction of this inadequacy will require more than just bringing pressure to bear on the workers for better work. Assuming a change in definition, changes in survey procedure may still be recommended." First, out of consideration of economy, it may be feasible to confine the canvassing of cities and towns to a sample but retain the complete coverage of the open county zone (Table 5). Second, require the investigator to call on *all households* in his district (except those in towns and cities which are not in his sample) and make out a report for each—whether the household operates a "farm" or not. The identification of farms would be made in the central office where a consistent policy could be established and kept under control. This procedure is proposed to eliminate or reduce unknown incompleteness in return for one which has a calculable amount of sampling fluctuation. The complete census of all households in the open country whether farm or non-farm would probably have considerable use to market researchers, students of population problems, etc., who now must depend on data obtained on a somewhat unsteady and not too well-known base.

Inadequate definition of operation (*farmer*). In general the word *operator* (when dealing with farms) is used to describe *a person* who "operates" a farm. If two persons jointly operate a farm both would generally be regarded as operators and each partner would so regard himself. Complete independence in the operation of a farm is therefore not required before the status of "operator" can be reached. It may be helpful to define some concepts on this matter as follows: Let the entrepreneurial function required of each farm be performed by a person or a number of persons which collectively will be called the *operatorship*. Each farm must have one and only one operatorship but that operatorship may consist of one or more persons called operators. Now an operator may be associated with more than one farm through membership in more than one operatorship so the total number of operators for a given area could be a different figure than the total number of farms for that area.

A distinction of this sort is not made by the Bureau of the Census. To quote its instructions, "For a farm operated by two or more partners enter *only one of the partners* as the operator, preferably the senior partner, unless the junior partner is actually conducting the operations." (Italics mine.) This definition of operator, for the sake of simplicity, does violence to reality.

UNIVERSITY OF MICHIGAN LIBRARIES

Forcing multi-person operatorships into this mold brings about some curious results. There will be an undercount of the number of true operators—that is, the kind which the economists, sociologists and the laymen would be thinking of. Likewise a number of other undercounts result such as number of farm dwellings having electricity, etc. (only one dwelling from each farm is eligible for this information), mortgage debt (asked of the one operator only and of only those who own land), land owned by the operator personally (here again asked only of the one operator), etc. This difficulty could be overcome by adopting a more complex model than now used—one permitting of multi-person operatorships.

Inadequate definition of agriculture. Although the Bureau of the Census calls its quinquennial survey a "census of agriculture" there is some doubt that it is comprehensive. This "census of agriculture," for example "covers" about 60% of the land area of the United States (Table 6). There may be some objection to the regarding of the remaining 40% as containing no "agriculture". For example, land under forest

TABLE 6

TOTAL AREA IN LAND AND IN FARMS, U.S., BY AGRICULTURAL CENSUS YEARS 1920, 1945[a]

(1)	(2)	(3)	(4)	(5)
Year	Total Land Area (Acres)	Total Area in Farms		Total Farms (Number)
		(Acres)	(Per cent)	
1945	1,905,361,920	1,141,613,510	59.9	5,859,169
1940	1,905,361,920	1,060,852,374	55.7	6,096,799
1935	1,903,216,640	1,054,515,111	55.4	6,812,350
1930	1,903,216,640	986,771,016	51.8	6,288,648
1925	1,903,216,640	924,319,352	48.6	6,371,640
1920	1,903,215,360	955,883,715	50.2	6,448,343
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.

[a] Data from published reports of the Bureau of the Census.

is regarded by the Bureau of the Census as a farm enterprise (that is one on which information is obtained) if it is *on* a "farm"; if such land is not on a farm it is not reported in the "agricultural census." Land under grass is similarly treated—for example "land used under a grazing permit is not to be included" as part of the areal extent of a farm. Even land that contributes directly to the production of food such as "production of maple sirup or sugar with no agricultural operations"

and "picking or gathering of wild nuts, wild fruits, or wild plants (medicinal, ornamental, etc.) except where the land is maintained primarily for their production" is not reported in the agricultural census "unless there are *agricultural* operations". (Italics mine.)

"Farming, or agricultural operations", according to the Bureau, "Consists of the production of crops or plants, vines and trees (excluding forestry operations) or of the keeping, grazing, or feeding of livestock for animal products (including serums), animal increase, or value increase". This seems like a reasonable definition of *agriculture*—except for the exclusion of forestry operations—but by attempting to make this concept synonymous with *farming* the Bureau has made some rather curious decisions, some of which have just been mentioned. The resulting census appears to be one of *farming* rather than of *agriculture*, and should be so labelled. Or, as an alternative, the scope could be broadened to include *non-farming* agriculture[7] as well. By this means the statistical no-man's land, comprising 40% of all United States land, can be brought under adequate statistical description. By bringing these two segments under one comprehensive periodic census it will be possible to eliminate the unknown duplications, omissions and other difficulties to which the present arrangement is subject when the Forest Service, Bureau of Agricultural Economics and other agencies attempt independently to get information on the neglected 40%.

Inadequate unit for presentation of data for small areas. Although it does not publish its agricultural census data on geographic areas smaller than a county, the Bureau of the Census will provide data by Minor Civil Division if and when someone is willing to pay for it. The smallest unit on which the Bureau can provide data is the enumeration district, which is usually the Minor Civil Division (township, beat, etc.) or some subdivision thereof. The Minor Civil Division is of course a unit determined primarily for political purposes. In many states its boundaries are difficult to determine and are short-lived. They do not necessarily describe areas having local significance such as a community.

There is a great need for agricultural data on a small area basis which can be effectively used for land use experimentation, research and planning. There is not only a need for data for areas smaller than a county but also for areas larger than a county but smaller than a state, e.g., the irrigated valleys of the west.

It is suggested that for agriculture a "natural" area be adopted in-

[7] And include all animal and plant production including fish and a number of other categories presently omitted by the Bureau of the Census.

UNIVERSITY OF MICHIGAN LIBRARIES

stead of the Minor Civil Division. This area might be based on that which has a common drainage, such as a valley. These will be permanent and will have local significance. They could be the bricks out of which valley authority areas could easily be constructed. They would not interfere with statistics for counties but would greatly facilitate the compilation of useful valley data.

Summary of proposals. A number of inadequacies in the census of agriculture have been discussed, using the 1945 census as the main illustration. Some of these were traced to difficulties of concept such as the conflicts between that which is *agricultural* and that which is *farming* and the operator as a *person* versus the operatorship as an *economic function*. Improvement can be made here by proper labelling and by broadening the scope.

To improve the completeness of enumeration it is suggested that the questionnaire be constructed suitable for use with all households whether they be associated with farms or not. In the open country all households should be canvassed but in cities and towns the canvassing should be confined to a sample (in the agricultural census taken above) in order to avoid heavy costs.

It is suggested that in place of the Minor Civil Division a natural area based on drainage be adopted as the basic unit for providing information on small areas.

THE EDGE MARKING OF STATISTICAL CARDS

A. M. LESTER
Montreal

THE SYSTEM OF edge-marking of statistical cards described by Dr. Thurstone in the September issue of the Journal has an additional application of considerable importance which he did not mention. This was discovered in the work of analyzing flying accident causes during the recent war. The main data for each accident could be expressed in from 100 to 200 "Yes-No" answers. The total number of accidents to be dealt with was sufficient to make a fully mechanical punch-card system a reasonable possibility, but it was found that reference to a brief written history of each accident was so frequently needed during analysis that hand-sorting of cards which could contain such histories was superior to full mechanization. Cards were utilized with edge punching and the needle was used for most sorting operations, particularly, of course, for replacing the cards in their filing arrangement.

The usual objections to the punching and pinning method were found, however, i.e. errors of punching were difficult to repair satisfactorily and virtually involved making out a new card each time; the edges of the cards eventually tended to tear after much use, and to pin through several thousand cards in search of an element occurring only a very few times was a laborious procedure which might be unreliable owing to cards sticking in the pack.

A system of marking the rims of the cards instead of punching the edges was therefore tried, meaning the actual thin rims of the cards and not the one-eighth of an inch of the front nearest that rim. Fairly thin cards were being used which ran at about 100 to the inch, and it was found not only that a tiny mark on the rim of one of them made with the back of a pen using India ink or red ink was easily visible (when the cards were carefully blocked so that their rims presented an even surface) but also that quite small deviations in the position of the marks could be at once detected. The rims on some types of card were too spongy and caused the ink to run as blotting paper would, but with good quality cards and a fine pen a mark could be made quickly and accurately.

This form of rim marking does not have the advantage of Dr. Thurstone's method, that both sides of the card can be used. On the other hand it can claim the following points in its favor as compared with Dr. Thurstone's scheme, or as compared with the ordinary edge punching scheme:

UNIVERSITY OF MICHIGAN LIBRARIES

(1) *ease of reference*—a count can be made while the cards are in their filing drawers if they are tapped into alignment to ensure accuracy;
(2) *ease of extraction*—cards can be picked out without disarranging the other cards. For many types of analysis this is a particularly valuable asset;
(3) *ease of correcting errors*—ink marks on the rims of cards can be bleached or whitened over just as they can on the front of a card but they also can be shaved off with a sharp pair of scissors without seriously impairing the trim of the cards;
(4) marking of the rims of the cards can be used in conjunction with edge punching and with Dr. Thurstone's type of edge-marking.

This type of rim marking might be of value in many statistical analyses although, of course, it has its limitations. It is particularly applicable where, as in accident statistics, a large number of different elements may occur only a few times each in a large universe. In this type of statistical analysis one of the most frequent questions to be answered is how often a certain element occurred in several thousand instances, and what were the circumstances in each case. If the occurrence of the element in question has been marked on the rim of the cards, it is a matter of a few minutes to run through the drawers and pick out the cards in question for further study. For this type of analytical procedure, rim marking would often be quicker than mechanical sorting from punched cards, as well as being enormously cheaper.

CONRAD ALEXANDER VERRIJN STUART (1865–1948)

Another notable statist, honorary member for thirty-five years of the American Statistical Association and outstanding member for fifty years of the International Statistical Institute, has become a part of the history of his major field of work. His friend and compatriot, H. S. Methorst, who succeeded him as Secretary General of the Institute in 1911 and held that post for a quarter of a century, will write for its Bulletin an obituary note based on a wealth of information personal and professional far greater than mine.

It remains for me to outline those aspects of Verrijn Stuart's life which will interest especially the readers of this Journal and accounts of which have been printed in other languages than Dutch. I knew him as a fellow member of the Institute for half a century during which period he missed not one of its twenty biennial sessions (1899–1938); in those sessions he played a leading role. In 1899 he became director of the Central Statistical Bureau of the Netherlands the reorganization of which he described at that time in the Institute's Bulletin and some twenty years later for our Association in Koren's "History of Statistics." In 1902 he printed in the Bulletin a paper on Birth Rates, Still Births and Infant Mortality in certain Dutch Cities and County Districts. Nine years later as Secretary General of the Institute and Chairman of the Committee which organized its Thirteenth Session he welcomed its members and guests to the Netherlands and The Hague.

After the first part of the second Thirty Years War he presented to the Institute as its Rome session (1925) a note on the Representative (Sampling) Method and five years later at the Tokio session a Report on National Capital and Income in the Netherlands. In the following year he submitted to the Financial Committee of the League of Nations a Memorandum on the Gold Question.

In the Revue of the International Statistical Institute which began in 1933 he published a paper (1935) on the Causes of Death and another (1938) on Cancer in the Netherlands. Of more permanent interest probably are his reports on the Mexico City (1933), London (1934) and Athens (1936) sessions of the Institute. The suggestion which he made and repeated in those reports looking towards an improvement in the scientific character of the sessions grew out of an experience almost unparalleled in length and depth, during four years of which he had been Secretary General and then for twenty-seven years an independent friend and critic of his successors, a suggestion

in which I heartily concurred. He thought that the work of the Institute would be improved by a drastic reduction in the number of topics considered at a session, a reduction in the number and importance of section meetings, and a concentration on a few subjects to be probed in the general sessions.

This suggestion was in line with changes made at the Washington session in 1947, changes far more radical than those to which he had looked forward ten years before. We may hope indeed that the forward steps in the progress of international statistics then taken will be found a generation hence to be comparable in importance with those taken between 1875 and 1890 when the Institute arose out of the ashes left by the Franco-Prussian War which had killed its predecessor, the International Statistical Congress. It was for some such an advance that Verrijn Stuart worked and hoped.

WALTER F. WILLCOX

PROCEEDINGS

108TH ANNUAL MEETING

HOTEL STATLER, CLEVELAND, OHIO

MINUTES OF THE ANNUAL BUSINESS MEETING

The American Statistical Association convened for its 108th Annual Business Meeting on the evening of December 28, 1948, at the Hotel Statler in Cleveland, Ohio. A motion was made and passed to approve the minutes of the last Annual Business Meeting held in New York at the Hotel Commodore, December 29, 1947.

Aryness Joy Wickens, Chairman of the 1948 Committee on Fellows reported concerning the fellows elected by the Committee.*

George W. Snedecor announced the appointment of W. Edwards Deming as the new member of the Committee on Fellows to serve for a period of five years from January 1949 through December 1953. Merrill M. Flood gave the reports of the Secretary and Treasurer.*

Walter A. Shewhart, one of the retiring Directors of the Association read the report of the Board of Directors on activities for 1948.* This report had been approved by the Board at its meeting on Monday, December 27, 1948.

Isador Lubin announced that the ASA and the American Economic Association would participate jointly in a memorial service for Wesley C. Mitchell to be held at the Hotel Cleveland, Wednesday, December 29, 1948.

Joseph Berkson reported approval of the activities of the Biometrics Section during 1948.

Isador Lubin read the minutes of the Commission on Statistical Standards and Organization and asked that the recommendations be presented to the membership for such action as they might wish to take.

In the discussion following the motion to accept Mr. Lubin's report, Theodore Brown stated that he thought the report opened the way for, or even by interpretation directed, the Statistical Association to investigate any situation in which it did not approve of the statistical results. Mr. Brown said that he had not the slightest objection to a Commission which would set up standards of good quality in statistical methodology or of ethics in their use, but that in general he believed the work of the Commission should be strictly limited to the setting of such standards.

He indicated that the Commission might well act as an arbitrator in a situation in which a dispute had arisen involving either statistical procedures or their interpretation, but the Association participation, if wise, would be limited to those cases in which the request for a review of a situation originated outside of the Association.

He took very definite exception to the report which seemed to him, from the single reading of Mr. Lubin, to imply that the Association might, on its own initiative, attempt to investigate any statistical procedure in private industry which in the opinion of the committee, or the members representing the Association, or of the Association itself was deemed to be bad statistical practice.

---

* See the report of the Committee in this issue of the *Journal*.

He stated that, "Such police investigations into the work or actions of private business cannot be carried out within the American system of free enterprise and might react so as to bring serious harm to the Association."

Phillip J. Rulon moved that Mr. Lubin's report be accepted without implication that its recommendations be carried out. In further discussion of the motion Harold Hotelling requested clarification of what ASA action would result in "regimentation of private enterprise." Isador Lubin read the following excerpt from the annual membership meeting at Atlantic City in 1947. "The report of the Committee on Standards was approved after some discussion and recommendations that the standing committee to be appointed be expanded to include subcommittees of Association Fellows who would work on the development of standards." Mr. Brown maintained that while he agreed with the original vote of 1947, he did not agree with the proposed recommendations.

Milton Epstein proposed an amendment to the motion which would provide that the words "without implication" be striken from the motion. Theodore H. Brown moved that the entire matter be referred to the Council instead. The Epstein motion was put to the vote and was carried by a substantial majority. Morris M. Copeland moved that the motion be amended to provide that the report be referred to the Council without recommendation. This motion was carried by a standing vote.

President Snedecor expressed his great appreciation of the work of the various committees carried on during 1948, and expressed his particular gratitude to the Committee on Nominations for its excellent work in preparing the slate of officers.

Gertrude M. Cox, Chairman of the Committee on Elections, read the official report of the Committee.*

Simon Kuznets, newly-elected President of the Association, accepted the chair, and as his first official act, gave the floor to George W. Snedecor, retiring president, who delivered his presidential address.**

President Kuznets expressed the gratitude of all members of the Association to George Snedecor for his active leadership during 1948.

Merrill M. Flood announced that the 1949 Annual Meeting would be held in New York, December 27–30, 1949.

Helen M. Walker, reporting for a Resolutions Committee consisting of herself and Morris M. Copeland, presented the following four resolutions:

RESOLVED: That the officers and members of the American Statistical Association express to Gale Ober and the members of his committee their sincere appreciation for the careful planning of all the arrangements for these meetings.

RESOLVED: That the officers and members of the ASA express deep appreciation for the excellent program prepared by members of the Program Committee, namely: Joseph Berkson, Ernest Blanche, John Cover, A. Ford Hinrichs, Simon Kuznets, Rensis Likert, Abraham Wald, Allen Wallis, Harry Wellman, Aryness Joy Wickens, and Merrill M. Flood, *Chairman*.

Whereas the Placement Committee of the New York Chapter has, during the past year, carried on a remarkably effective service to its members, RESOLVED: That the Board and Council give serious thought to ways in which such placement service can be more adequately financed and can be extended on a nationwide scale.

---

* See Report of the Committee, in this issue of the *Journal*.

** March, 1949, *Journal of the ASA*.

Whereas those members of the ASA who have major interest in the social sciences constitute a large proportion of the present membership of the Association, and whereas many of these members have expressed dissatisfaction with the benefits received from their membership, RESOLVED: That the Board and Council give serious thought to ways and means of providing a balanced program of services to all its members.

These four resolutions were voted on by the membership and approved.

The following two resolutions were transmitted by Miss Walker without any recommendation from her Committee.

Resolution on Marriage and Divorce Statistics: Whereas, the American Statistical Association recognizes the need for more adequate marriage and divorce statistics, to provide data needed in statistical research in many areas such as demography, health, business, welfare and others; and

Whereas the development of national vital statistics of marriages and divorces depends upon cooperative state-federal relationships along the lines which have proven effective for the vital statistics of births and deaths:

RESOLVED, that the American Statistical Association recognizes the need for state centralization of marriage and divorce records and statistics and for their integration with other vital records and vital statistics; and urges all states not yet operating such an integrated system to initiate such a program at the earliest possible time; and

RESOLVED, that the American Statistical Association calls upon the Federal Security Agency, through its National Office of Vital Statistics in the Public Health Service, to encourage the development of systems of marriage and divorce records and statistics in every state leading to maximum comparability and prompt availability of data; and to undertake the development of detailed national marriage and divorce statistics, adequate to meet the pressing needs for such vital statistics.

Resolution on Birth Statistics: Whereas, birth statistics are essential to developing local, state, and national estimates of population changes, determining trends in family size, and are required for planning and evaluating health, welfare, and related programs, and

Whereas, the usefulness of birth statistics would be increased by improved registration completeness and by a current knowledge of the completeness in which births are registered in local areas and by various characteristics of the population, and

Whereas, such a knowledge would also provide the basis for promotion of complete registration,

RESOLVED, that the American Statistical Association strongly urges that a uniform, nation-wide test of registration completeness be carried out in conjunction with the 1950 decennial census of population in order to provide the required measures of registration completeness on local, state, and national levels and that every effort be made to obtain more complete registration.

The membership voted to refer these two resolutions to the Council for consideration.

Aryness Joy Wickens rose to present the following resolution.

RESOLVED, that a special Sub-Committee be set up by the Committee on Committees to consider and formulate specific standards for the selection of Fellows, in carrying out the constitutional provision that fellows shall be "statisticians of established reputation"; that the recommendations of this special

Sub-Committee be forwarded to the Board and Council for its approval; whereupon the approved standards shall constitute the criteria for the selection of fellows.

Theodore H. Brown, Robert W. Burgess, and Stuart A. Rice spoke for the resolution. W. Edwards Deming took the position that the Committee on Fellows should review its own standards. The resolution was put to the vote and was passed in the form in which Mrs. Wickens presented it.

Eugene Pike introduced a resolution which would provide that the Council, where it is in the interest of the membership, state those matters of policy before it and ask the membership to communicate with the Council concerning it. Milton Epstein, in discussing the Pike resolution, asked that one-half day of each Annual Meeting be set aside for business meetings so that the membership might have an opportunity to prepare resolutions and discuss all matters of policy before the Association. The resolution was voted as presented. Helen M. Walker reminded the assembled members at this point that annual business meetings are not mandatory under the new Constitution and that any such provision would have to be made on the initiative of the officers of the Association.

The meeting was adjourned.

Report of the Board of Directors

The year 1948 has been one of transition in the affairs of the Association. It has been marked by the orderly movement toward the organization structure envisaged in the new Constitution, by the reorganization of the national office under a new Secretary-Treasurer, and by an unavoidable increase in membership dues.

The new Constitution was adopted by mail ballot of the membership in February 1948 and became effective on January 1, 1949. Nominations and elections of officers for 1949 were in accordance with the new Constitution. For the first time in the history of the Association the membership voted by mail ballot for national officers. The representatives of the newly formed Districts were also elected to the Council. The Council, now the policy forming body of the Association, held its first meeting on December 29, 1948.

The 108th Annual Meeting of the Association was held in Cleveland between December 27 and 29. Joint sessions were held with the American Economic Association, the American Marketing Association, the Econometric Society, the American Farm Economic Association, the Population Association of America, the American Public Health Association, the Institute of Mathematical Statistics, the Biometric Society, and the Ohio Section of the American Society for Quality Control. Over 600 people attended the sessions.

It is noteworthy improvement that the planning of the 1948 Annual Meeting program was completed before October. Moreover, the 1949 Annual Meeting program has already been outlined, approved by the 1948 Program Committee, and referred to the 1949 Program Committee for consideration and further planning.

Definite arrangements have been concluded for a four-day meeting in New York City during the Christmas holidays of 1949. The fortunate circumstance that both the American Association for the Advancement of Science and the Allied Social Science Associations will be meeting in New York at that time made it possible for the Association to plan combined sessions that will cover an exceptionally broad subject field. Joint sessions have already been scheduled

with the American Economic Association, American Farm Economic Association, American Marketing Association, American Psychological Association, American Society for Testing Materials, Biometric Society, Econometric Society, Institute of Mathematical Statistics, Population Association of America, and the Housing Research Committee of the Social Science Research Council. There are tentative plans for other joint sessions with the American Society of Mechanical Engineers, American Astronomical Society, and the Psychometric Society.

The situation with respect to Association membershp is not entirely heartening. It will be recalled that the Board of Directors reluctantly recommended a substantial increase in membership dues from $5.00 to $8.00 per annum, the increase being financially unavoidable in order to meet continually increasing costs of printing and other services in an inflationary period. This decision was ratified by mail ballot of the membership in February 1948. The withdrawal of a portion of the membership in the face of this dues increase, either because of a marginal interest in statistics or because their real incomes were adversely affected by the inflation, had, of course, been anticipated and discounted in advance. The actual result for 1948 is that the size of the Association has remained about constant. About 800 members had dropped out and an about equal number of new members have come in. While this withdrawal of 800 members is regrettable, it is surely not a recurrent item. During the year, new membership applications totaled approximately 725 as against 1,100 in 1947. This slackening of the rate of increase can probably be attributed to the deterrent effect of the higher dues in this and other Associations, and to the fact that vigorous canvassing in 1946 and 1947 of the most likely sources of new members has forced recourse to more marginal groups from the standpoint of their concern with statistics.

The main organizational change in the national office has been the appointment of a new Secretary-Treasurer after the resignation of Dr. Lester S. Kellogg from that position at the 1947 Annual Meeting. Sylvia C. Weyl took over as Acting Secretary-Treasurer on an ad-interim basis until May. During the spring of 1948 an ad hoc nominating committee, under Samuel S. Wilks, energetically searched the field for a man combining the many managerial and scientific attainments necessary for the executive leadership of the Association. After considering a number of candidates Merrill M. Flood was chosen, and his nomination as Secretary-Treasurer was unanimously approved by the Board of Directors. In addition to his duties as Secretary-Treasurer, Dr. Flood served as Chairman of the 1948 Program Committee which organized the 1948 Annual Meeting and planned the sessions to be held in 1949. The Council of the Association at its December 29 meeting re-elected Dr. Flood for a 3-year term.

Two new chapters were added to the Association in 1948. A Saint Louis Chapter is now functioning regularly, and the old and established Sacramento Statistical Society has become a chapter of the Association.

A temporary Committee on Publications was appointed by President Snedecor, with William G. Cochran as Chairman to outline the policies to be followed by the various Association publications, to suggest better means of cooperation among them, and to make other such recommendations within this general area as it deemed proper. This Committee has now tendered its report to the Board of Directors. Under the new Constitution, the standing Committee on Publications will continue this function.

The Commission on Statistical Standards and Organization, under the chairmanship of Isador Lubin, was organized and held its first meeting in November.

Three of its members participated in the survey by the Social Science Research Council of public opinion polling methods—with special reference to the inaccurate forecasts of the 1948 election returns. Several problems of major importance have been placed before the Commission. It will surely be in a position of influence and develop rapidly during 1949 as a positive force for the advancement of statistics.

With the 1950 Census of the United States already approaching, the Census Advisory Committee has been actively engaged in aiding the Director of the Census in reviewing and appraising schedules. It has screened many requests from the public to the Census Bureau for inclusion of additional questionnaire items. This Committee has an added responsibility now that sampling methods are under fire and yet of such critical importance in the work of the Bureau of the Census.

Within the field of Association publications, the Biometrics Bulletin has been increased to four 64 page issues, and both format and contents have been greatly improved. The name of the publication has been changed to Biometrics. The American Statistician, launched in the summer of 1947, has been very well received by the membership and has already established itself as a valuable periodical of the Association.

The Association cooperated informally in the work of the National Bureau of Economic Research under Frederick C. Mills, for the Hoover Commission on the Organization of the Executive Branch of the Government. At the request of Dr. Mills, President Snedecor arranged for an independent study of the statistical activities of the Federal Government and preparation of a report for the member group. John W. Tukey canvassed a representative group of members of the Association, many of whom are familiar in detail with the statistical practices and problems of the Federal Government, and obtained their evaluation of the caliber of federal statistical work and their recommendations on its consolidation, extension and improvement. His excellent report, which summarizes the judgments obtained was forwarded to Dr. Mills.

The International Statistical Institute advised the ASA of the adoption of its new statutes, which provide for the affiliation of national and international statistical societies. The ISI has been reorganized to work as an active professional organization dedicated to the furtherance of statistical science throughout the world. In view of the similarity of purpose of the Institute and the Association, Stuart A. Rice, president of the ISI, expressed the hope that the Association would request affiliation. The proposal is being considered by the Council.

The Association has an increasing interest in the applications of statistics to management problems, many of its members are active in such fields as business economics, marketing, and quality control. The National Management Council, a federation of national organizations concerned with management problems has invited the Association to become a member. Among its present members are the Society for the Advancement of Management and the American Society of Mechanical Engineers. The invitation will be acted upon by the Council.

The Eastern North American Region of the Biometric Society requested affiliation with the ASA during the past year. Decision on this application will be made by the Council. The Biometric Society, an international organization of biometrics workers, decided in 1948 to designate Biometrics as its official organ. Block subscriptions to Biometrics were made available to members of the society at a reduced rate as a means of assistance for the new Society.

Plans for several new activities of the Association have been partially de-

veloped during the year. Among these are establishment of a speaker's bureau to serve the chapters, formation of a research group to study the program and operations of the Association, and intiation of regional District meetings. Some or all of these new projects will be activated during 1949. These new projects are an integral part of your Board's program during 1949.

Regrettably, the need for economy during 1949 made it necessary to postpone indefinitely the project for launching the periodical "Statistical Reviews."

The past year has been one of unavoidable financial stress and structural reorganization. The Board considers that the major problems involved in this postwar reorganization have been faced and are now on the road to solution. The task ahead of the Association is to consolidate organizationally and financially and at the same time to continue to expand its membership and carefully selected activities, in the face of an inflationary trend that is prejudicial both to professional societies and to the memberships they serve.

GEORGE W. SNEDECOR, *President*
ISADOR LUBIN
LOWELL J. REED
WALTER A. SHEWHART
FREDERICK F. STEPHAN
SAMUEL STOUFFER
WILLARD L. THORP
MERRILL M. FLOOD, *Secretary-Treasurer*

The Secretary's Report on Membership

There were 4,720 members remaining at the beginning of 1948 after 189 were dropped for non-payment of dues. Although 725 new members joined during 1948, and 27 were reinstated, there was a net decrease of 70 bringing the total on December 31, 1948 to 4,650.

The 1948 membership is composed of the following groups:

Honorary members	13
Fellows	138
Student members	290
Regular members	4,209
Total Membership	4,650
Corporate Members	6

Over 200 members resigned during the year and some 600 were dropped for non-payment of dues. The primary cause for this heavy loss in membership is undoubtedly the increase in dues from $5.00 to $8.00. The unfortunate delay in settling the matter of dues increase, and the consequent lateness in getting dues notices to the members, also had an adverse effect on membership.

The members of the Biometrics Section, at the end of December 1948, number 960, of whom 170 are associate members and 790 are also regular members of the Association.

The deaths of the following members were recorded during the year: Wesley C. Mitchell, *Fellow;* Edward G. Benson, Bertram Butler, Ruth Dawson, Valentino Dore, T. Bertrand Graham, Edward W. Higgins, Arthur Hurd, Walter E. Magney, Douglas W. Oberdorfer, Hallie K. Price, O. A. Pope, H. M. Tompkins, John H. Watkins, *Regular Members.*

MERRILL M. FLOOD, *Secretary*

Report of the Nominating Committee

The Nominating Committee of the American Statistical Association announces the election of the following officers for the year 1949.

President	Simon Kuznets
President Elect	S. S. Wilks
Vice President	
3 years	Dorothy S. Brady
2 years	H. A. Freeman
1 year	Lester S. Kellogg
Directors (3 year term)	
	C. H. Goulden
	L. L. Thurstone
District Representatives	
Western District	Maurice I. Gershenson
	Henry B. Moore
North Central	Howard L. Jones
Southeastern	Samuel Weiss
	Morris H. Hansen

They were selected as persons

(1) who are distinguished in their contributions in applying methods to problems in various fields
(2) who are acquainted with and contributing to statistical theory
(3) who represent the various fields of membership interests and the different geographical regions of the country.

Respectfully submitted,
Nominating Committee:
MILTON FRIEDMAN
CHARLES F. SARLE
MORTIMER SPIEGELMAN
HOLBROOK WORKING
GERTRUDE M. COX, *Chairman*

Report of the Committee on Fellows

The American Statistical Association has announced the election of four of its members as Fellows—three American statisticians and one Canadian statistician. (There are now 145 Fellows of the American Statistical Association from a total membership of about 4,700.)

The newly elected Fellows are:

Eugene L. Grant, Professor of Civil Engineering at Stanford University, who has developed special courses for teaching statistics for engineers, and who has been a leader in the practical application of statistics to engineering.

Tjalling J. Koopmans, of the Cowles Commission for Research in Economics at the University of Chicago, who has done outstanding work in the application of statistics to economics.

W. Allen Wallis, Professor of Statistics and Business Economics at the School of Business at the University of Chicago, who was Administrative Director of the Statistical Research Group of the Office of Scientific Research and Development which developed and supplied answers to many statistical problems of military importance submitted by the armed services.

J. W. Hopkins, Canadian biometrician, of the Division of Biology and Agriculture of the National Research Laboratory of Canada and Coordinator of the Special Committee on Applied Mathematical Statistics, of the National Research Council of Canada. His principal work has been done in applying statistical methods to experiments with grains and livestock. During the war he served as scientific advisor to the Air Vice Marshal of Canadian anti-submarine warfare.

MINUTES OF THE MEETING OF THE COMMISSION ON STATISTICAL STANDARDS AND ORGANIZATION

Present: S. Wilks
W. Shewhart
F. Croxton
I. Lubin

Absent: L. Reed

CHARTER RECOMMENDATIONS

The Commission recommends that the Committee on Committees, in preparing the charter for the Commission on Standards, should define its functions and competence in the terms stated in the report approved by the members of the American Statistical Association at its annual business meeting on January 25, 1947, as follows:

> The Committee should have rotating membership based on a three year term; one-third of the initial membership to be appointed for one year, one-third for two years, and one-third for three years. Members should be eligible to reappointment. Election should be made by the Board of Directors after consultation with the Committee.
>
> The functions of the Commission should be:
>
> A— to provide a tribunal to render opinions and recommendations on controversial issues relating to statistical procedure and presentation of statistical material.
>
> B—develop a list of minimum standards for published statistical materials.
>
> C—upon request from governmental bodies, review actual or proposed undertakings and make recommendations relative to standards.
>
> The Committee might eventually develop a code of ethical practices in statistical work.

It is recommended that these terms be amplified to the extent of adding at the end of the last sentence above the words "with a view to enhancing the status and acceptance of statisticians by the general public."

Rules of Procedure. The Commission decided to make no recommendations relative to rules of procedure. It was the consensus that these rules should be developed as experience dictates.

Membership. The Commission recommends that Joseph Davis of Stanford University and Samuel Stouffer of Harvard University be added to its membership.

Budget and Staff. It is impossible at this time to make any definite estimate as to the budgetary requirements of the Commission. Its financial requirements will depend upon the amount of work that will be undertaken during the coming year, the nature of the projects it undertakes and the extent to which Governmental agencies will finance studies and surveys made upon Governmental request.

The Commission, however, feels that it is essential that there be a part-time secretary available to it. Eventually it will be necessary to have a full-time secretary. In order to finance a part-time secretary and the expenses of Commission and Sub-Commission members, it is estimated that for the year 1949 approximately $15,000 will be required if the Committee is to undertake projects that are listed below.

Projects

The Commission gave consideration to possible projects that it might undertake in the immediate future.

1. *Joint Committee on the Economic Report.* The Commission had before it a request submitted by the Staff Director of the Joint Committee on the Economic Report of the House and the Senate relative to its reports on "Statistical Gaps" and "Economic Indicators." The Commission felt that this request should be complied with and that it should undertake to study and assess both of these Joint Committee reports. It is proposed to set up a Sub-Commission at an early date to undertake this project.

2. *Consumer Price Index.* The Commission feels that a request from the Commissioner of Labor Statistics to advise the Bureau of Labor Statistics concerning technical questions involved in constructing the Consumer's Price Index be complied with. It was the feeling of the Commission, however, that the success of projects of this sort will be greatly enhanced if its services were requested before any definite action has been taken by the agency involved. In other words, the Commission would prefer that it be called in at the time that technical problems are in process of being studied rather than to pass judgment at a later date upon already determined procedures.

3. *Standards for Publication of Federal Statistical Data.* The Commission agreed that it should comply with the request of the Bureau of the Budget for a review of the Bureau's report on "Standards for the Publication of Statistical Data." It felt, however, that it was not within its competence to comply with the Bureau's request that it suggest principles that should guide the Federal Government in its release or suppression of statistical information during an emergency.

4. *The Adequacy of Present Types of Federal Statistics.* At the April 1948 meeting of the Board of Directors a resolution by Mr. Solomon Barkin to the effect that the American Statistical Association should investigate its responsibilities in the field of industrial statistics was referred to this Commission. At a later date the Labor Advisory Committee of the Bureau of the Budget adopted a resolution submitted by Mr. Barkin requesting that the ASA be called upon to establish a committee to appraise the adequacy of present types of statistical data and the direction in which statistical information should be developed.

The Commission considered these resolutions and is of the opinion that no action should be taken upon them until the report of the Hoover Commission is made public. Considerable time and energy has been devoted by the Staff of the Hoover Commission to the question of the collection and development of federal statistics. Indeed, a paper on the work of the Hoover Commission in the field of statistics will be delivered at the forthcoming annual meeting of the American Statistical Association by Dr. Mills. The Commission recommends that any possible action be deferred until the Hoover Report can be studied and analyzed.

5. *The Kinsey Report.* It has been suggested that the Commission review the Kinsey report with a view to passing judgment on the methodological statistical

techniques used by the author. It was the unanimous opinion of the Commission that we do nothing in this field unless specifically requested by a representative body of American citizens.

6. *Polling Techniques.* The widespread criticism of the election polling results has resulted in requests that the Commission initiate studies of the techniques used by the various polling organizations. In view of the fact that the Social Science Research Council has appointed a special committee to study the election polls and in view of the further fact that two members of this Commission are on the SSRC Committee, it was felt that no action should be taken in this field at this time. The Commission was aware of the fact that further work in the field of polling would have to be done after the SSRC Committee completes its report. However, provision has already been made for such work by the National Research Council and the Social Science Research Council. Arrangements have been made for studying special problems such as sampling, interviewing, and panels. It may be advisable after these studies have been completed for the Commission to look further into the question of polling techniques. At the moment, however, it feels that anything it might undertake would be more or less in the nature of duplicating work already under way.

Independent Investigations. The Commission feels that in those instances where it is not specifically requested by a Government agency, by a foundation or similar body which financed or sponsored a study, or by a representative body of citizens to undertake investigations of projects or of statistical standards it should act on its own initiative only in those instances where the public interest, domestic or international, is concerned. As problems arise that are affected with the public interest, the Commission will take under consideration the advisibility of investigating them with a view to pointing out to the public such weaknesses and inadequacies of statistical techniques and presentation as it may find. Its primary purpose shall be to further higher standards in statistical fields and thereby enhance the status of statisticians in the mind of the public.

Report of the Treasurer

The 1948 budget as originally approved by the Board of Directors planned for an income of $61,900 and expenses of $58,925. Actual income was $51,320 and expenses were $64,605. In September, the Board voted a revised budget providing for the increase in expenses and decrease in income.

Income was less than expected primarily because some 600 more members than had been estimated left the Association during the year. Also there were some 275 fewer new members than estimated, and receipts from subscriptions, sales, and advertising were some $4,000 less than estimated. Income for 1948 exceeded that for 1947 by $14,315, nevertheless, because of the additional revenue derived from higher dues and subscription rates.

Expenses were somewhat higher than anticipated because of increases in printing costs. Printing costs will be still higher in 1949.

The national office plans to continue its drives for new members and to make a new drive for additional advertising and subscriptions. Through this expansion the Association will widen its sphere of influence and increase its services, and at the same time improve its financial status.

MERRILL M. FLOOD, *Treasurer*

Report of the Auditors

To the Board of Directors of
American Statistical Association

We have examined the attached financial statements of American Statistical Association relating to the year ended December 31, 1948. Our examination was made in accordance with generally accepted auditing standards, and accordingly included such tests of the accounting records and such other auditing procedures as we considered necessary in the circumstances.

The recorded cash receipts for the year were traced to the deposits shown on the bank statements and the amounts for dues and subscriptions were tested with the membership and subscription records. The paid checks were inspected and related vouchers tested in support of cash disbursements for the year. The bank balances were reconciled with amounts reported direct to us by the depositaries and the cash on hand and the securities owned at December 31,1948 were verified by inspection. We did not check the membership and subscription records in detail or make any independent verification of the inventory of old Journals, the office records of which are based, in part, on data assembled in prior years, no recent physical inventory having been taken.

The life membership reserve at December 31, 1948 reflects the amount needed to support a life annuity for each life member in the same annual amount as that which could have been purchased by the original lump sum payment, based on a 2¼% interest rate, the 1937 Standard Annuity Table and the age of the life member when the lump sum payment was made, in accordance with a resolution of the Board of Directors adopted pursuant to a mail ballot in January 1949. Previously the reserve had been calculated on the basis of the combined annuity table of mortality with assumed interest at 4% per annum and an assumed annuity of $5 per member. The amount treated as income from life memberships in 1948 represents the excess of the reserve at the beginning of the year over the required reserve, on the new basis, at the end of the year.

The accounts for the year 1948 include for the first time a provision for employees' accrued annual leave. The reserve of $1,209.54 provided at December 31, 1948 includes $506.50 applicable to the prior year. We understand that this provision was made at the direction of the Secretary.

In our opinion, the accompanying statements present fairly the position of American Statistical Association at December 31, 1948 and the results of its operations for the year, in conformity with generally accepted accounting principles applied on a basis consistent, except as mentioned in the two preceding paragraphs, with that of the preceding year.

PRICE, WATERHOUSE & Co

Washington, D. C.
April 27, 1949

AMERICAN STATISTICAL ASSOCIATION BALANCE SHEET

Assets	*December 31, 1948*	*1947*
Cash in bank and on hand	$ 7,513.00	$ 1,853.91
Accounts receivable	1,204.97	1,356.22
Investments:		
United States Savings Bonds, Series D, at redemption value	6,406.00	6,138.00
Stocks, at cost (at market quotations $8,219 and $6,150, respectively)	5,793.50	5,793.50
Inventory of old Journals, at approximate cost	1,907.51	2,024.63
Furniture and equipment, at cost less depreciation	2,892.19	2,482.76
Deferred charges	277.80	
	$25,994.97	$19,649.02
Liabilities		
Accounts payable	$ 4,631.43	$ 6,515.76
Note payable	5,000.00	
Accrued interest	37.50	
Accrued annual leave	1,209.54	
Deferred income (collections applicable to subsequent year):		
Dues	15,964.53	1,054.50
Subscriptions	3,797.59	2,704.76
Other		115.35
	$30,640.59	$10,390.37
Net worth:		
Life membership reserve	$ 3,131.40	$3,750.20
Surplus, per statement	(7,777.02)	5,508.45
	$(4,645.62)	$ 9,258.65
	$25,994.97	$19,649.02

SURPLUS STATEMENT	*Year ending December 31, 1948*	*1947*
Balance at beginning of year	$ 5,508.45	$ 4,248.79
Add—Transfer of balance in Centenary Sustaining Fund		6,379.51
	$ 5,508.45	$10,628.30
Deduct—Excess of expenses over income for the year, per income statement	13,285.47	5,119.85
Balance at end of year	$(7,777.02)	$ 5,508.45

American Statistical Association
Income Statement

	Year ending December 31,	
	1948	*1947*
Income:		
Dues—current year	$ 32,495.90	$21,298.58
Dues—prior years	54.65	215.00
Life membership income	418.80	397.33
Subscriptions	7,386.28	6,206.95
Advertising	1,728.77	2,010.08
Reprints	892.09	509.00
Journal sales	2,702.19	2,827.80
Biometrics Section income	4,212.70	2,466.86
Miscellaneous	724.15	684.70
Dividends and interest	704.50	398.57
	$ 51,320.03	$37,014.87
Expenses:		
Journal—printing, mailing and reprints	$ 12,160.70	$10,105.81
Salaries and wages, including in 1948 $506.50 accrued annual leave expense applicable to prior year	30,582.81	17,037.26
American Statistician Bulletin	5,875.55	4,259.88
Biometrics Section expenses	4,316.29	2,366.46
Rent	2,880.00	1,365.00
Office supplies, printing and mimeographing	1,599.27	1,764.42
Postage	1,748.74	1,511.11
Telephone and telegraph	754.81	512.03
Travel expense—officers	649.20	603.91
Annual meeting expense	613.31	647.89
Depreciation of furniture and equipment	442.86	324.99
Promotion expense	743.28	
Loss on disposal of equipment		151.09
Storage of old Journals	71.95	72.00
Cost of old Journals sold	573.10	382.27
Miscellaneous	1,593.63	1,030.60
	$ 64,605.50	$42,134.72
Balance, loss, carried to surplus	$(13,285.47)	$(5,119.85)

BOOK REVIEWS

Edited by

OSCAR KRISEN BUROS

Rutgers University

Actuarial Statistics: Vol. II, Construction of Mortality and Other Tables. *J. L. Anderson* (Scottish Widows' Fund and Life Assurance Society, 9 St. Andrew Square, Edinburgh, Scotland) and *J. B. Dow* (Standard Life Assurance Co., 3 George St., Edinburgh, Scotland). Published for the Institute of Actuaries and the Faculty of Actuaries. London N.W. 1: Cambridge University Press (Bentley House, 200 Euston Road), 1948. Pp. xvi, 281. 21*s*.

REVIEW by T. N. E. GREVILLE

Chief, Actuarial Analysis Branch, Public Health Service
Federal Security Agency, Washington 25, D. C.

THIS book has been written primarily to assist British actuarial students in their preparation for the examinations of the Institute of Actuaries and the Faculty of Actuaries. As might be expected, the main emphasis is on the construction of mortality tables from the records of insured lives and life annuitants. However, the two chapters on the construction of national life tables at least acquaint the reader with the more important problems that usually arise in that connection; and the final chapter on sickness rates is also very informative to the uninitiated.

The principal topic, the construction of mortality tables from the records of insured lives, is rather fully discussed, and there is no doubt that this is the most complete treatment of the subject thus far published. The very detailed explanation of exposed-to-risk formulas will be especially helpful. Throughout the book the exposition is lucid and accurate. Clarity is assisted by numerous illustrative examples, which add greatly to the usefulness of this text. The absence of problems to be worked out by the student is a shortcoming which it has in common with other British actuarial textbooks.

It is natural that the principal emphasis should be placed on developments in Great Britain. However, after making all due allowance, it is still the impression of the reviewer that the book suffers somewhat from a certain insularity of approach. Certain useful contributions have been made by actuaries in the United States and Canada which, it is suggested, could most appropriately have been mentioned in a book purporting to deal comprehensively with a branch of scientific methodology. One is surprised, for example, to find no mention of abridged processes of life table construction, a subject in which there have been important developments on this side of the Atlantic, but in which British demographers have expressed considerable interest. Again, on page 255, where it is pointed out that occupational mor-

tality statistics based on national population data "are not reliable guides for a life office to use for assessing extra premiums for occupation," some mention might have been made of the extensive studies of occupational mortality among assured lives in the United States and Canada made by the Actuarial Society of America and the Association of Life Insurance Medical Directors.

This is, of course, a minor criticism; and it is the reviewer's opinion that the authors have, on the whole, handled their material in expert fashion and have produced a text which will be most useful not only to the students for whom it is intended but also to any others having an interest in the subject matter.

Experimental Designs in Sociological Research. *F. Stuart Chapin* (Professor of Sociology, Chairman of the Department, and Director of the School of Social Work, University of Minnesota, Minneapolis, Minn.). New York 16: Harper & Brothers (49 East 33rd St.), 1947. Pp. xi, 206. $3.00.

REVIEW BY MARGARET JARMAN HAGOOD
Statistician, Division of Farm Population and Rural Life
Bureau of Agricultural Economics, Washington 25, D. C.

IN THIS brief volume, Dr. Chapin has summarized the designs and findings of nine sociological "experiments" and, in addition, has presented a considerable amount of general exposition of experimentation in sociology and of sociometric scales. As Chapin points out, the work is complementary to, rather than in competition with, Ernest Greenwood's *Experimental Sociology*, which appeared two years earlier. Greenwood treated the subject with primary attention to the conceptual formulation and logical principles involved, with a critical examination of the literature. Chapin, on the other hand, had as his purpose to "illustrate the method of experimental design by reproducing concrete studies" and to provide "a source book of examples of specific application analyzed in some detail" (p. ix).

Dr. Chapin has fulfilled this purpose excellently—in fact, no other person in the United States is in position to fulfill it equally well. To Chapin goes the credit for leadership among sociologists for trying to adapt the experimental approach to sociological research problems and for decades of work and encouragement of work of others in this field. He and his students have made outstanding advances in quantification of hitherto unmeasured social phenomena and in continuously carrying out pioneering work to establish sociology as a science.

Without reflection on the author, one can differ with him, as to when the term "experiment" or "experimental" is to be validly used. I am inclined to be much more restricted than is Chapin when it comes to classifying a social research project as an "experiment." The social "experiments" reported in his book do not measure up to the criteria that I hold for "experiments."

But I do agree with Chapin that in our field, the research worker is unrealistic who writes as though one could control the "treatments," and provide for random selection of the units which are to receive "treatments," as in the case of biology and in some areas of psychology.

R. A. Fisher's methods of analysis of variance and covariance have not yet been fully explored in comparison with Chapin's matching technique which sacrifices much data. Dr. Chapin is aware of this and it is hoped that he will focus research on the problem in the future.

Elements of Mathematical Statistics. *C. V. L. Charlier.* Including **Table of Poisson's Function** by *L. v. Bortkiewicz.* Edited and translated by *J. A. Greenwood.* Brooklyn 25, N. Y.: J. A. Greenwood (25 Winthrop St.), 1947. *Two reviews follow:*

REVIEW BY BURTON H. CAMP
Emeritus Professor of Mathematics, Wesleyan University
Middletown, Connecticut

THIS well-known book, written in German and published in Hamburg in 1920, is now available in an English translation, and it is this translation which is the subject of this review. The translator has also made occasional emendations of the text and an improvement in the tables. He himself has computed new tables for ϕ_2, ϕ_3 and ϕ_5, these functions being derivatives of the function defining the normal curve; he has made some corrections in the original tables; and has added a new table, No. 44. The meanings of Tables 43 and 44 are not indicated in the appendix where the tables occur. It can be learned from the preface that Table 43 is the table of Bortkiewicz, that is, a table of the Poisson function, the reader being left to infer the meanings that must be attached to the numbers at the top and side of the table. There is no easy way of finding out what Table 44 does mean, but the thorough reader will find it explained in a footnote on page 60.

So far as the reviewer can judge the translation is sufficiently accurate. The English is often awkward but not without a quaint charm so that perhaps by its very defects the phraseology makes one the more aware of the originality of the distinguished author who is thus being presented to us. In fact one would read Charlier's book now more to become acquainted with Charlier's way of looking at things rather than to study the material which the book contains. This material has been worked over many times since 1920 and has been presented in various forms in several books; but there is much value nevertheless in going to the original source, and this translation makes it possible for the English reader to do this. The general scope of Charlier's book may be indicated as follows. First of all there is a discussion of homograde or alternative statistics. These are statistics showing the number of times an event occurs, in a given number of trials, when it may either occur or not occur. This sort of frequency distribution may be

approximated by successive terms of the point binomial, in certain cases, and in other cases by the modifications of the point binomial introduced by Poisson and by Lexis. Heterograde statistics comprise frequency distributions that are now sometimes called continuous. These may be approximated sometimes by the normal probability curve, sometimes by the more general A-Type or B-Type curve. The subject of simple correlation is considered at some length, and the discussion includes what Pearson calls tetrachoric correlation.

REVIEW BY ALEXANDER M. MOOD
Statistician, Project RAND, Douglas Aircraft Co., Inc.
1500 Fourth Street, Santa Monica, California

IT IS always interesting to read an older book and compare the state of statistics then with now. The present little volume was written in 1910 (the translation was made from a 1920 German translation) just two years after Student's famous paper was written and before the full import of that paper was recognized. Fisher's paper on the foundations of statistics was not to appear for ten more years.

The book is divided into two parts of about fifty pages each; the first part discusses what would now be called discrete distributions, and the second, continuous distributions. The discussion is largely in terms of the first four moments of the distributions and of course in the second part there is considerable attention to the Gram-Charlier system of frequency functions. Actually there is much in the book that can be found in the best sellers in statistics today, but unfortunately that fact cannot be interpreted as a compliment to Charlier's sagacity—it is merely an indictment of our best sellers. I refer to such things as the detailed computational instructions in terms of class frequencies all outmoded by modern computing machines, the hoary errors about skewness and kurtosis still fed our students today, the confusion between population parameters and their estimates, estimation by the method of moments only, the great faith in the standard error whether the distribution is normal or not. But Charlier, at least, cannot be much blamed for these shortcomings because they were either not relevant or had not been pointed out at the time.

The real trouble with the book is simply that it was written before the Fisherian revolution in statistics. It was written in that period when statisticians were trying to develop and formalize statistical inference in terms of relatively simple classes of frequency functions—the Pearson curves and the Gram-Charlier series. Looking back, it is easy enough to see that such a course was actually a blind alley, that a fitted curve contains no more information than the data to which it is fitted. But at the time that approach must have looked promising and Charlier did significant work in the field.

Today the book is of little interest except possibly to students of the history of statistics, and the reviewer is at a loss to rationalize the presentation of this translation.

Elements of Nomography. *Raymond D. Douglass* (Professor of Mathematics) and *Douglas P. Adams* (Assistant Professor of Graphics). (Massachusetts Institute of Technology, Cambridge, Mass.) New York 18: McGraw-Hill Book Co., Inc. (330 West 42nd St.), 1947. Pp. ix, 209. $3.50. [London W.C. 2: McGraw-Hill Publishing Co., Ltd. (Aldwych House, Aldwych), 1948. 21*s*.]

REVIEW BY JOSEPH ZUBIN
Associate Research Psychologist, New York State Psychiatric Institute
722 West 168th St., New York, N. Y.

THIS is an elementary text of nomography for students in engineering which may serve very well as an introductory text for any student of statistics who wishes to know how nomographs may be constructed. From the point of view of the statistician it is unfortunate that no direct applications of nomography to statistical problems are made. Perhaps that is a task for a future nomographer who has statistics well in hand. This simple elementary text ought to prove useful for those students who wish to become acquainted with the use of nomographs in their field.

The more routine tasks of computation such as comparison of percentages and comparison of correlation coefficients with their standard errors are more readily accomplished by means of nomographs than by calculating machines and tables. There are several source books in which such nomographs are provided but the manner of their construction and limitations on their use is not always understood by the average statistician. Furthermore, no new statistical aids of this type can be invented unless statisticians become aware of the problems involved in nomograph construction and of the areas in which they may prove useful. The simplicity of the present text and its empirical approach ought to win adherents for the use of nomographs on a wider scale.

A Guide to Public Opinion Polls, Second Edition. *George Gallup* (Director, American Institute of Public Opinion, Princeton, N. J.). Princeton, N. J.: Princeton University Press, 1948. Pp. xxiv, 117. $2.50. [London E.C. 4: Oxford University Press (Amen House, Warwick Square). 14*s*.]

REVIEW BY ROBERT COBB MYERS
Educational Testing Service, Princeton, N. J.

THIS is a revised edition of the question-and-answer handbook which was first prepared in 1944. The avowed purpose of the book is "to answer, in non-technical language, the questions that people most frequently ask about public opinion polling." If *non-technical* can be considered as synonymous with *naive, unsophisticated, incomplete, repetitious,* and *contradictory,* we can then concede that the author's purpose in this respect has been fulfilled.

Many popular misconceptions to the contrary, neither the American Institute of Public Opinion nor the publisher, the Princeton University Press, is a part of Princeton University. Each, however, is bound to the university

by many ties. Although the AIPO is a privately owned commercial organization, its director is listed in the Princeton University catalogue as a member of the Advisory Councils of the Psychology Department and of the School of Public and International Affairs. The University's Office of Public Opinion Research serves as an official depository and reference guide for the end products of all AIPO, or Gallup, surveys—these archiving activities being under the direction of Princeton's Professor Hadley Cantril, and financed by the Rockefeller Foundation. In addition, the book, *Gauging Public Opinion*, by Cantril and various of his graduate students, which was first published in 1944 by the Princeton University Press, carries a formal dedication to George Gallup. The situation is, therefore, understandably confused; and one is often at a loss to know just where Gallup and his AIPO leave off and where Princeton University begins. However, despite this confusion, if *A Guide to Public Opinion Polls* is properly regarded simply as a popular publication by a business owner-manager about one of his several enterprises then it is subject to fewer strictures than could be levelled against it if it were otherwise considered. There is, after all, little justification for criticizing a book for not being scholarly or objective when it was never intended by the author that it contain these qualities.

The book, having been published prior to last November's presidential election, presents answers to the author's self-posed questions in a most assured, if not supercilious, fashion. The curious reader will find no mention of an "infant science" which latterly has become so popular a catchword with the commercial pollsters and their apologists. Critics of Gallup's procedures, when mentioned at all, are dismissed as being irresponsible, uninformed or undemocratic. Their criticisms are rationalized away either by retreat to a reverent quotation from James Bryce or by some breezy *non sequitur*. For example, Kornhauser's (not mentioned by name) careful exposition and criticism of the antilabor bias of AIPO from 1940 through 1945 is met by this reply: "Carried to the extreme, it would have meant that every time survey results showed the public hostile to a Hitler, another survey had to be found which was favorable to him."

Gallup gives an earnest defense of his "quota" method of sampling as compared to "area" or "probability" sampling. He also gives much lip service to statistical theory and Bernouillian "laws of probability" in discussing applicable formulae for determining size of sample in relation to probable error; and he compliments Theodore Brown and Samuel S. Wilks for the help they have given pollsters by "their studies on the statistical problems involved in sampling." The lay reader might well gain the impression that a considerable body of statistical theory underlies Gallup's "quota" operations, but the inquiring statistician will be relieved to find this admission on page 30: "While the accuracy of quota sampling cannot be determined from mathematical formulae—principally because of the inability to calculate the interviewer selection factor—there does exist a growing record of performance of public opinion polls which have in the past been operating on quota sampling procedures."

An entirely new section is devoted to reprinting a paper entitled "The Quintamensional Plan of Question Design." This was first delivered by Gallup in London before a meeting of the executives of his foreign affiliated enterprises, and later published in the *Public Opinion Quarterly*. The space consumed by this is one-thirteenth of the entire text of the book. One wonders at its disproportionate inclusion among "questions people most frequently ask about public opinion polls." It is difficult to imagine a deluge of letters and telegrams inquiring: "What is the quintamensional plan of question design?" More pertinent, however, would be an answer to the question concerning what proportion of AIPO surveys on controversial issues ever make use of this five-point plan involving the use of (*a*) filter or information questions, (*b*) open or free answer questions, (*c*) dichotomous or specific issue questions, (*d*) reason why questions, and (*e*) questions bearing upon intensity of opinion. The book gives us no clue as to the extent to which this laudable plan is actually used. On the surface it would seem too expensive for more than token use in a commercial operation.

A considerable compendium might be composed of important questions regarding opinion polling which are neither posed nor answered in this book. Some of these unanswered questions have become far more pressing than heretofore in view of the rent through the clouds which occurred last November, and many are currently being answered. But an extended discussion of these matters is not particularly apposite in a review.

The title is as misleading as that given to the first Kinsey report. Not only is this not a guide to opinion polls; it is not even a guide to all of Gallup's own polling activities—under whatever corporate names. Unfortunately for such organizations as Chicago's NORC, Michigan's Survey Research Center, and Columbia's Bureau of Applied Social Research, the public is all too likely to take the title at its face value.

If you want to learn in the shortest possible time the most complimentary things that Gallup has to say about the operations of his American Institute of Public Opinion, then this is the book for you.

The Decomposition of a Series of Observations Composed of a Trend, a Periodic Movement, and a Stochastic Variable. *A. Hald* (Assistant Professor, University of Copenhagen, Copenhagen, Denmark). Copenhagen, Denmark: G. E. C. Gads Forlag, 1948. Pp. 134. Paper.

REVIEW BY
D. B. DELURY, *Director, Department of Mathematical Statistics*
Ontario Research Foundation, Toronto 5, Canada
AND
BOYD HARSHBARGER, *Professor of Statistics*
Virginia Polytechnic Institute, Blacksburg, Virginia

THE TITLE of this work indicates clearly its content. The development concerns a series of observations $y_{\nu i}(\nu = 1, 2 \cdots n;\ i = 1, 2 \cdots k)$, with measurements of a concomitant variable $x_{\nu i}$ and assumes an underlying

model of the form

$$y_{ri} = \Gamma(x_{ri}) + \eta_i + \epsilon_{ia},$$

where Γ denotes a trend, η a periodic element and ϵ a random component. The trend is assumed to be representable by a polynomial.

The fitting of this model to data is accomplished by constructing orthogonal polynomials in x (for the case in which the recorded x-values are in arithmetic progression and the y's have equal weights), such that, after elimination of the η's, the normal equations have diagonal form. Thus, the trend may be fitted one term at a time until a "satisfactory" fit is reached.

These orthogonal polynomials are, of necessity, considerably more complicated than those appropriate to fitting a trend only. Each pair of (n, k) values requires it own set of polynomials. Hence a table of values of these polynomials must be much more elaborate than that given, for example, in Fisher and Yates' tables. Part III of this study contains tabulated values of the polynomials up to degree 5 for the following (n, k) pairs.

$k = 3$	$n = 5, 7$
$k = 4$	$n = 25$
$k = 5$	$n = 5, 7, 8, 9, 10, 11$
$k = 12$	$n = 10$

From the several fields in which this mathematical model might conceivably be appropriate, the author selects two for special mention, the treatment of economic time series and the analysis of agricultural field experiments. The decision on the usefulness of the model in dealing with economic series must be left to economists. Its value in the design and analysis of field experiments will surely be widely questioned.

The agricultural experiment chosen to illustrate an application of these methods is called a "row experiment," in which the plots are placed consecutively in a row and divided into blocks, so that each treatment occurs once within each block. The arrangement of treatments within blocks is the same for all blocks. The x's then refer to the positions of the plots and the η's are treatment parameters.

One could, as the author remarks, employ this model when the treatments are arranged in any manner whatever, but when the "periodic" arrangement is abandoned, orthogonal polynomials are no longer feasible. In this case, regression analysis (presumably using ordinary orthogonal polynomials) or, what amounts to the same thing, a covariance analysis with the position of the plot as concomitant variable, permits the use of the same mathematical model. Even when the treatments are randomly arranged within blocks, control of the error resulting from variation in fertility is possible by these means and such procedures have, in fact, been used a number of times. It is true that the computational labor is considerably greater when orthogonal

polynomials, of the sort developed in this book, are not available, but then, when random arrangements are used, there is good reason to expect that such refinements will not be needed and that a simple analysis of variance will suffice. On the other hand, if one adopts a periodic arrangement of treatments in the hope that fertility variations are expressible in terms of a polynomial of low degree, his computational program is dismal indeed if this hope is not fulfilled. Furthermore, it is not clear to the reviewer that randomization can be dispensed with without assurance that variation in fertility is completely accounted for by the fitted polynomial (a viewpoint probably held also by the author, see p. 91), but presumably it never is.

Whether or not this mathematical model is suitable for routine use in field experiments, it seems likely that there are many situations in which it is quite appropriate and a table of orthogonal polynomials to facilitate the computations it entails is a useful item of statistical equipment. It is to be hoped that a more complete table than the one included in this book will soon be issued.

Part I, devoted to theory, includes a survey of standard regression theory and applies these classical methods to the problem at hand. The requisite orthogonal polynomials are developed and the necessary distribution theory is derived. The method of moving averages is discussed with respect to data to which this mathematical model is applicable and some comparison is made between this method, the type of regression analysis here developed and the analysis of variance as applied, for example, to a randomized block experiment. The methods used in this section are direct and the exposition is clear and concise.

Part II presents three worked examples, the fitting of a time series, the analysis of a row experiment and a re-working of the data of an industrial example reported by H. E. Daniels.*

The trend in the time series is fitted to degree 6 without reaching a good fit. The polynomial representing fertility variation in the row experiment is terminated at the fourth degree, on the ground that the coefficient of the fifth degree polynomial is not significant—a questionable criterion for general use. The same reason is given for choosing a third degree polynomial in the industrial example. Regression analysis in this last case comes out at about the same place as does the analysis of variance used by Daniels. Some advantage is claimed for the regression approach, in that only 3 degrees of freedom are needed to accomplish what 24 are used for in the analysis of variance.

Part II contains also a short but excellent discussion of some of the theoretical considerations on which experimental design is based.

This work is well put together and written in a concise and straightforward style that makes for easy reading. It can be recommended to those who (*a*) would like to have a concise, accurate statement of the methods and

* "Some Problems of Statistical Interest in Wool Research." *J Royal Stat Soc* 5:89–112 '38.

principal results of regression theory; (*b*) are interested in the theory of moving averages; (*c*) may find appropriate the mathematical model to which the book is devoted (e.g., in engineering research); (*d*) are perplexed about the old controversy over systematic and random arrangements, even though the concept of randomness is not dealt with as discerningly as one might wish.

Quality Control Methods. *Clifford W. Kennedy* (Quality Control Engineer, Federal Products Co., 1144 Eddy St., Providence, R. I.). New York 5: Prentice-Hall, Inc. (70 Fifth Ave.), 1948. Pp. vii, 243. $4.75, trade edition; $3.55, text edition. *Two reviews follow:*

REVIEW BY SEBASTIAN B. LITTAUER
Associate Professor of Industrial Engineering
Columbia University, New York 27, N. Y.

FOLLOWING the extensive as well as intensive application of statistical quality control during the war, the appearance of a number of texts at a variety of levels of interest and with a diversity of emphasis was to be expected. Of the dozen or so of books which have now been published, *Quality Control Methods* seems to be the first which is admittedly addressed to shop operating personnel, and therefore is to be viewed in that light. One might, on that account, reasonably have anticipated a greater emphasis on operating procedures, practical examples, administrative and economic problems, and in general the important nonstatistical aspects of quality control which have yet to be adequately presented. The fact, however, is that this work is devoted primarily to reporting statistical aspects of quality control, supposedly to fill the gaps left by other authors' omission of "basic details and primary concepts."

The role of statistics in quality control is introduced by a general discussion of sampling and the presentation of a variety of sampling plans including some dozen pages on sequential sampling. Following a discussion of "batch control" in Part 2, some elements of statistics are taken up in Part 3. Control charts for variables are considered in Part 4, and the closing section is devoted to administrative problems.

Statistical quality control can be motivated naturally by introducing the principles of acceptance sampling which can, again, lead naturally into control chart procedures. This has been effectively done by another author. In the present work, however, acceptance sampling procedures are presented apparently because "the progressive business house establishes a receiving system designed to make *certain* that the goods delivered fulfill requirements before the invoice is paid." The presentation of the various sampling plans is not prefaced by any exposition of statistical principles nor is their practical use validated by bringing in the role of statistical control in acceptance sampling practices. The exposition is necessarily descriptive and follows from

one plan to another, single, double and sequential, with little foundation that can be called statistical. It seems questionable, at best, to devote so much space to an attempted simplified explanation which "follows closely the instructions in *Sequential Analysis of Statistical Data: Applications*" when the readily available S.R.G. work does the job so well and so thoroughly.

The present treatment neglects to show the natural place of acceptance sampling in statistical quality control viewed both as a system of thought and as a system of practices. The author seems to confuse estimation with testing hypotheses. For example, early in Part 1, page 12, the author advises: "Throughout the use of sampling methods, never lose sight of the fact that basically you are attempting to estimate or judge the whole by what you see and think from examining a part or a sample." While this is altogether wholesome advice it does not apply to the use of the sampling plans presented, and, in fact, the originators of these plans would argue quite strongly against this practice. In a number of other places the author makes reference to estimating either the "condition" of the lot, or the quality of the lot.

Many statements made by the author must be read with great care in order to avoid misinterpretation. On reading (p. 7) that, "A sample is defined by the dictionary as a part shown to prove the quality of the whole," this reviewer checked some six dictionaries. The closest approach to the quoted statement was found to be, "A part of anything presented as evidence of the quality of the whole," from *Webster's Collegiate Dictionary* (Merriam and Co., 1917). Perhaps a statistical dictionary is needed. Again (p. 31), "The sample size n may be a number of ounces or pounds, a proportion of a fifty-ton gondola of ore, for example, or it may be so many drops of liquid, or 100 persons interviewed during a public opinion poll in a certain city" does not say what it may have been intended to mean. Nor can the statement (p. 68) "In other words, the chances would be 95 per cent good (with the buyers' risk from error in sampling, p, at 5 per cent) that the lot was of quality equal to $p_1(p_1 = .05)$, or better" be regarded as correctly interpreting the use of a Wald sequential plan. In numerous other places in Part 1 the text requires discriminating reading that it is not likely to get from shop personnel who have had no previous statistical training. The whole treatment of acceptance sampling might have been clarified and simplified by use of OC curves. As it stands, the exposition is misleading to the uninitiated.

The remainder of the book is conventional except for the use of shop language. In the author's attempt to by-pass the preciseness of statistical inference for the suggestiveness of intuitive insight he has inadvertently exposed his whole exposition to question. Thus there is never a clear distinction between parameters and statistics, which leads to confusion between control limits and confidence intervals in the discussion of p-charts. This may also account for the absence of a clear-cut explanation of and emphasis

upon the concept of statistical control, stressing the importance of time order and rational subgrouping. There are no exercises for the student to work on and few adequate illustrative examples. In spite of the six practical problems in the last chapter, one might have hoped that this author would have embellished his writing with more material taken from his extensive experience.

It is rather surprising to this reviewer that there is no reference in the body of the text or in the index to Shewhart's contributions, and only a belated reference to his earlier book in the brief bibliography. In spite of all the space (26 pages) devoted to sequential sampling, Wald's name does not appear anywhere. On page 102, there is a replica of a table from page 23 of *Introduction to Industrial Statistics and Quality Control* by Paul Peach, with no specific acknowledgment. It is customary to acknowledge any complete reproduction apart from a general reference in the bibliography. Is it not the responsibility of the publisher to supply careful editing in order to pick up omissions like those mentioned.

REVIEW BY CHARLES R. SCOTT, JR.
Superintendent, Plant No. 1, SKF Industries, Inc.
Front St. and Erie Ave., Philadelphia 34, Pa.

THIS is an informative book about industrial quality control. The author handles the subject in a way that a beginner can understand. Rather than attempting to write a handbook putting forth a series of tables to be used in a mechanical sort of way, the author attempts to enable the beginner to secure a reasonable grasp of the statistical theory underlying industrial quality control. A discussion of the methods used in modern industrial plants carries the reader along through the theoretical sections of the book. The text is directed toward the practical man interested in statistical application of quality control at the operating level.

The book is divided into five sections. Part I discusses acceptance sampling, running through the gamut of sampling tables, sampling size, and the various accepted sampling methods. Part II discusses batch control and the practical methods of securing control by attribute inspection. Part III discusses and outlines practical uses of the frequency distribution and the standard deviation. In Part IV, Kennedy discusses average and range and compares the standard deviation with the average-range method. Of great importance is the way the author explains how to read a chart and interpret the data collected. Part V very nicely winds up with some good commonsense suggestions on what to do with the knowledge gained from the book. The reader is not left confused as to how to apply practically what he learned. The commercial examples form a basis for practical application. The book includes a bibliography, an appendix, and an index.

Metoder att Uppskatta Noggrannheten vid Linje- och Provytetaxering. [Methods of Estimating the Accuracy of Line and Sample Plot Surveys.] *Bertil Matérn* (Student in the Institute of Mathematical Statistics, Stockholm, and Research Statistician in the State Forest Research Institute, Stockholm). Stockholm, Sweden: State Forest Research Institute, 1947. Pp. 138. Paper.

REVIEW BY T. W. ANDERSON
Assistant Professor of Mathematical Statistics
Columbia University, New York 27, N. Y.

THIS book consists of a study of the use of information in a survey of a certain type for the estimation of the accuracy of that survey. This kind of systematic sample survey of a region is used, for example, in estimating certain characteristics of forests. The author utilizes the theory of stochastic processes in his treatment of the use of quadratic forms in observed variables for the estimation of the variances of estimates of various characteristics. The mathematical results are applied to a variety of problems arising in forestry surveys.

The kind of survey to which most attention is paid is a completely systematic sample. A specified grid of survey lines, say q, is placed in a specified way on the region Q, which is to be surveyed. The average of the characteristic on the grid lines, say $f(q)$, is the estimate of $f(Q)$, the average of the characteristic over the region. For instance, the estimate of the proportion of the area of Q which is forested is the proportion of the lengths of the survey lines running through forests. It is customary to estimate the "variance" of the error, $f(q)-f(Q)$, by means of a quadratic form in the values of the characteristic on pieces of the survey lines. The principal topic of this investigation is the selection of a quadratic form which will give a good estimate of the variance of the error regardless of the topographical variation.

The author suggests that a stationary stochastic process defined over a plane be used as a probability model for this problem. To each point (u, v) of the plane we attach a random variable $f(u, v)$ such that the set of random variables has the properties that the expectation of $f(u, v)$ is m and the variance is σ^2 and that the correlation of $f(u_1, v_1)$ and $f(u_2, v_2)$ is $\rho(t)$, where t is the distance between (u_1, v_1) and (u_2, v_2). If $\rho(t)$ is assumed continuous, we can define a random variable attached to a region as the integral of $f(u, v)$ over the region (with convergence meaning limit in the mean) divided by the area of the region. For example, $f(Q)$ is the integral of $f(u, v)$ over Q divided by the area of Q. The random variable associated with a line is a line integral of $f(u, v)$ divided by the length of the line. The expected value of a certain kind of quadratic form in such variables can be written as

$$\sigma^2 \int_0^\infty \rho(t)a(t)dt,$$

where $a(t)$ depends on the regions or lines involved.

The accuracy of a survey is indicated by the expected value $E\{T\}$, where $T = L[f(q) - f(Q)]^2$ and L is the length of q. The author finds an approximation to $E\{T\}$ and an associated "distance function" $\bar{a}(t)$. The expected values of different quadratic forms used to estimate $E\{T\}$ are studied by comparing the associated distance functions with $\bar{a}(t)$. Certain quadratic forms are suggested as best because their expected values are near $E\{T\}$ for all $\rho(t)$ in a certain class.

The remainder of the book deals with some questions of the relation of the model to reality and with the application of the theory to practical problems. Plot surveys, sampling trees on the lines and double-sampling schemes are also considered. A number of examples are given, and computational procedures are occasionally suggested. Although the main text is in Swedish, there is an extensive summary in English (20 pages), and the table and figure headings are given in both languages.

This study is of interest to statisticians on two counts, firstly as a study of stochastic processes from the point of view of statistical application and secondly as a presentation of means of choosing methods of estimating the accuracy of surveys. Of course, between the theory and the application is the interpretation of the probability model. A line survey here is a completely systematic survey; there is no chance mechanism involved in placing the lines. To enable him to apply probability calculus the author considers the formation of the topography under the grid lines as a stochastic process. The combined effects of Nature and man in giving the region the characteristics to be surveyed is taken to be that of a process defined by this model. Such an interpretation of the model leads to results different from those that would be obtained by considering placing a grid at random on a region with a given topography.

Except for one feature of this study, this approach seems to give results that have reasonable interpretations for the practical problems. In fact, answers are given to questions which have previously been answered inadequately or not at all. In devoting the remainder of the review to the one questionable feature the reviewer does not wish to detract from the otherwise fine work.

The purpose of a survey q is to estimate some characteristic of a given region Q. One would like to say something about the discrepancy between the estimate, say $f(q)$, and the value for this specific region, say $f(Q)$. For example, if $f(q)$ were normally distributed with mean value $f(Q)$ and variance τ^2, one could state the inequality $f(q) - \tau t \leqq f(Q) \leqq f(q) + \tau t$ with confidence α if t were chosen so the integral of the standard normal density from $-t$ to t were α. Confidence α would be interpreted as being approximately the proportion of times such a statement would be correct if the statement were made for a number of situations (specified by $f(q_1)$, $f(Q_1)$, τ_1^2; $f(q_2)$, $f(Q_2)$, τ_2^2; . . .). α would be independent of how $f(Q)$ would be chosen (as long as the choice is independent of $f(q)$. In particular, if $f(Q)$ were fixed the above theory would hold.

In his treatment the author takes $f(Q)$ to be a random variable, not a parameter. Thus, $E\{[f(q)-f(Q)]^2\}$ is an average over all possible topographical variations defined by the stochastic process over Q. But this is not the average which one would like to use to define the accuracy of a survey of a specific area. One would like the average taken over all topographical variations for which $f(Q)$ is the value that this region actually has. In short, one would like $f(Q)$ to be treated as a constant.

The probability model specifies a distribution of the set $f(u, v)$ which is the product of the distribution of $f(Q)$ and the conditional distribution of the set $f(u, v)$ given $f(Q)$. The use of the distribution of $f(Q)$ implies that we believe that $f(Q)$ for the particular region Q we survey arose in a way described by the model. Now this is unrealistic because we may have chosen Q to survey simply because we think $f(Q)$ is in a certain range of values. This question of using Bayes Theorem has been discussed so much that there is no need of doing more in a review than pointing out that this is what the author does.

It is possible to avoid this difficulty by using the conditional distribution of $f(u, v)$ given $f(Q)$. Then the question arises as to whether the mathematical results of this investigation hold for the conditional distribution. Since the author's conclusions are based on the distribution of $f(q_j)$ (values on line segments q_j) and $f(Q)$, we need only consider the conditional joint distribution of $f(q_j)$ given $f(Q)$. The author's conclusions rest on approximate equalities of the sort $E\{T_k\}=E\{T\}$, where T_k is a quadratic form in $f(q_j)$. Do these hold if $f(Q)$ is held fixed; that is, is $E\{T_k|f(Q)\}=E\{T|f(Q)\}$ true? It can be shown that in general the second equality does not follow from the first. This is easily demonstrated if the original process is normal because then the joint distribution of $f(q_j)$ and $f(Q)$ is normal. Thus, the actual conclusion we wish to draw from the theoretical study does not follow directly.

It is possible, however, that these results hold in some approximate sense. If the region Q is large and if $\rho(t)$ decreases rapidly, $f(Q)$ is distributed with a small variance, and, hence, may behave nearly like a constant. Then the effect of holding $f(Q)$ fixed may not be great. However, in order to justify his conclusions the author needs to study this question.

The author has exercised considerable ingenuity in solving difficult mathematical problems. The development is rigorous, except that many approximations are used; the effects of the approximations are studied to verify that they are legitimate. The problems of application considered seem to be important and interesting. It is to be hoped that the author in a future study will show that the objection to the use of Bayes Theorem approach can be met.

Lecture Notes on Mathematical Theory of Probability and Statistics. *Richard von Mises* (Gordon McKay Professor of Aerodynamics and Applied Mathematics, Harvard University, Cambridge, Mass.). Graduate School of Engineering, Special Publication No. 1. Cambridge 38, Mass.: Harvard Graduate School of Engineering (c/o Miss Natalie Nicholson, Librarian, Pierce Hall), 1947. Pp. 300. Paper, mimeographed. Out of print.

REVIEW BY BENJAMIN EPSTEIN
Associate Professor of Mathematics, Wayne University
Detroit 1, Michigan

THESE are notes on a year course in probability and statistics given for undergraduates and graduates at Harvard University. The author has, in the opinion of this reviewer, given an excellent presentation of some of the most fundamental ideas in probability and statistics. It is a pity that these notes are not generally available. They would represent a welcome addition in a field where the really good elementary texts can be counted on the fingers of one hand.

The frequency concept of probability is the dominant theme throughout the text and the author gives a very clear introduction to the basic notions underlying his approach to the axiomatic foundations of probability. In particular he defines at the outset the concepts of limiting frequency, chance as distinguished from randomness, collective, and the basic operations of place selection, mixing, partition, and combination.

There are many workers in probability and statistics who object to the way in which von Mises formulates his axioms, to the way in which formal and empirical aspects of probability are mixed.* Personally this reviewer prefers the treatment of statistics as given by Cramér in his recent book, *Mathematical Methods in Statistics*. However, I am of the opinion that it would be a mistake to make the preference so strong that one becomes blind to the originality and the basic character of the von Mises approach. While there appear to be grounds for preferring the measure-theoretic approach (which is of course at heart also a frequency theory) to that of von Mises, one should recall that von Mises was a pioneer in the field of probability and that he grappled, as far back as 1919, with many of the problems underlying the foundations of probability.

This reviewer, when reading the text, was made keenly aware more than once of the fact that the author is a master in the field and is a man who is not satisfied with a superficial glossing over of difficulties. He meets difficulties head on and brings them out into the light of day where they can be analyzed. This is particularly true in Chapters 3, 6, and 9. In these chapters he deals with (*a*) the weak law of large numbers; (*b*) Bayes' theorem and its relation to the problem of inference; (*c*) certain aspects of the Neyman-

* For an evaluation of this point, *see:*

1. Mises, R. von. "On the Foundations of Probability and Statistics." *Ann Math Stat* 12:191–205 Je '41.
2. Doob, J. L. "Probability as Measure." *Ann Math Stat* 12:206–14, Je '41.
3. Mises, R. von, and Doob, J. L. "Discussion of Papers on Probability Theory." *Ann Math Stat* 12:215–17 Je '41.

Pearson theory of testing hypotheses; (*d*) the contrast between Bayes' theorem and the Neyman-Pearson theory of testing hypotheses; and (*e*) the confidence interval concept. He shows very clearly the role played by the a priori distribution of the unknown parameter or parameters in determining the kind of probability inferences that one can draw from a sample. These are precisely the sort of things which should be brought to the attention of the student who is just entering the field of statistics.

The short sections on runs and Mendelian heredity are also very good and illustrate important statistical concepts. Another valuable feature of the book is that many of the problems given to the student are non-trivial and thought-provoking. The problems are drawn from many fields and help to illuminate parts of the theory that might otherwise have remained unclear to the student.

The treatment throughout is rigorous and the author does not hesitate to call on techniques in modern analysis which are about on the level of those given in Whittaker and Watson's *Modern Analysis*. Theorems are carefully stated and where concepts are involved the author takes the trouble to point out what would happen if certain underlying assumptions were not satisfied.

The selection of certain sections or chapters for special comment is not to be taken to mean that the rest of the book is not well presented. The reviewer has only picked out a few of the things that were treated *unusually well* in this book and that are either omitted or glossed over in most texts. Some of the other topics are taken up as well in a book such as Cramér's.

By way of summary, it is my opinion that the continual emphasis by von Mises on underlying concepts makes good sense scientifically and pedagogically. I also feel that this book gives the student some insight into the historical development of the subject and into the nature of the early difficulties and paradoxes. It is particularly the beginning student who needs to be reminded over and over again of just what assumptions are made when one makes a probability statement or an inference from data. It is refreshing to find an author who is honest enough to make the following explicit statement (Chap. 6, p. 16): "It remains an invariable fact, dominating all problems in mathematical statistics, that *no substantial inference can be drawn from a small number of observations if nothing is known 'a priori,' i.e., preliminary to the experiments, about the object of experimentation.*"

Statistics in School. *W. L. Sumner* (Senior Lecturer in Education, University College, Nottingham, England). Oxford, England: Basil Blackwell & Mott Ltd. (49 Broad St.), 1948. Pp. vii, 183. 9*s*. 6*d*.

REVIEW BY F. G. CORNELL
Professor of Education and Director
Bureau of Research and Service, College of Education
University of Illinois, Champaign, Illinois

THIS book, evidently intended for use as a text, is a summary of lectures given to postgraduate students in education at the University College, Nottingham. As such, it lacks some of the features which might be expected

of a book purposely prepared as a textbook. The author recognizes the difficulties of students of education with standard mathematical treatments on statistics. This is a problem of teacher training institutions not only in Great Britain, but also in the United States. Unfortunately, a successful formula has not yet appeared for the intuitive, genuinely nonmathematical, yet technically adequate and sufficiently comprehensive, treatment of statistics which would reach larger numbers of educational leaders and the rank and file of teachers.

The author of this book is reasonably successful in some of his attempts "to make it as simple as possible." Included are some topics needed and usually not found in educational statistics such as an introduction to analysis of variance and factorial analysis. With these, however, there are questions of the value of the limited depth of treatment to which the author restricts himself evidently to avoid complexity. To what extent can a "simplified" text compromise with the criterion of adequacy of content coverage, for instance, by including factorial analysis but limiting its discussion to tetrad differences? In his presentation of partial correlation, the author covers only the three-variable problem without multiple regression, error of estimate, and other concepts useful in understanding the subject.

The chapter on correlation and regression is typical of what is likely to result from confining the subject of educational statistics to 183 pages. In that chapter there appear technical errors of omission or commission which would lead an immature student to incomplete or incorrect information. Regression is introduced primarily as a means of defining the Pearson-r. Thus, only the deviation form of the regression equation is given. The essential concepts of "error of estimate" and "prediction" oddly appear under a section headed *Spearman's Footrule*, instead of in the section on "regression" where it appears to belong. In this chapter, some technical and sematic difficulties appear in such condensed statements as: "In (c) there is perfect correspondence between the scores and correlation is complete *and there is no regression.*" In education and psychology we usually find that correlation, if present, is *partial positive correlation.*" Incidentally in a later chapter the author gives the old probable error for r formula, $.6745\ (1-r^2)/\sqrt{N}$, and the Fisher z transformation as "hyperbolic arctangent," in a way which would not encourage use of the latter.

Certain important omissions prevent this book from being a significant addition to the field. Not included is a modern approach to sampling. A careful distinction is not made between large sample and small sample method, "statistic" and "parameter." Notions of sampling distribution are introduced through the medium of the "probable error" anachronism. Chi square is presented with the limited emphasis of "goodness of fit" and contingency.

The book throughout would be improved if there was greater use made of applications to educational problems and if examples were not excluded more or less to the rather sterile subject of school marks. The writing is heavily orientated in the direction of marking, to which relatively little

concern has been given in education in this country in recent years. A broader orientation of a book on school statistics seems essential if it is to contribute significantly to the study of modern educational problems.

Mathematical Treatment of the Results of Agricultural and Other Experiments, Second Edition. *M. J. van Uven* (Professor in the Agricultural University of Wageningen, Netherlands). Groningen-Batavia, Netherlands: P. Noordhoff F.V., 1946. Pp. viii, 310.

REVIEW BY G. A. BAKER
Assistant Professor of Mathematics and Assistant Statistician in the Experiment Station, College of Agriculture University of California, Davis, California

THE PRESENT tendency in the design and analysis of field trials seems to be to carry the elaboration of the mathematical models first devised by R. A. Fisher to greater and greater extremes of complication. The theory of many such models is well worked out and tables provided so that the probabilities of the errors in decision made in any trial can be estimated. The difficulty is that the Fisherian models, in many cases, do not sufficiently approach reality in their fundamental assumptions about the behavior of soil fertility. Professor van Uven is refreshingly careful and thorough in his fundamental assumptions about the behavior of soil fertility. He assumes, that soil fertility is a continuous function of position that changes slowly. His methods in this respect agree with Neyman's method of parabolic curves (1929). There is some evidence that, in some cases at least, this assumption of slowness of change in soil fertility is not sufficiently realistic.

Professor van Uven's tests of significance are based on estimates of mean differences and standard errors obtained by the methods of least-squares and asymptotic averaging. Reference is then made to the normal probability curve regardless of the sample size and character of the population distribution. The whole discussion of significance is naive as compared with those of J. Neyman, E. S. Pearson, and R. A. Fisher.

Professor van Uven discusses some of the simpler Fisherian designs in detail. In this respect he says that although sometimes the influences at work on the yield will not fully justify the hypotheses yet in many cases they will be very well applicable. This remark, of course, applies to situations with which Professor van Uven is familiar.

A list of technical terms in nine languages is included.

The second edition is a photographic reproduction of the first edition published in 1935. Professor van Uven intended to make some changes based on his further experience in teaching and examining field trials but could not because of shortages due to the war.

The book is well and logically written and proceeds on the basis of assumptions that are clearly stated. It will well repay study by anyone interested in the analysis of field trial data.

The Fourier Transforms of Probability Distributions. *Aurel Wintner* (Professor of Mathematics, Johns Hopkins University, Baltimore, Md.). Baltimore 18, Md.: the Author (Rowland Hall, Charles and 34th Sts.), 1947. Pp. v, 185. Paper. $3.00.

REVIEW BY J. WOLFOWITZ
Associate Professor of Mathematical Statistics
Columbia University, New York 27, N. Y.

THIS publication is a set of notes, taken by Dr. F. W. Light, of a course given by Professor Wintner in 1942–3. It begins with a description of a (one-dimensional) distribution function. Successive chapters are entitled: transforms, moments, convolutions, convergent convolutions, convergence in measure, semi-groups (i.e., infinitely divisible laws), inversions, examples, and projections (i.e., multidimensional and marginal distributions). Some references are given at the end of the book.

The reviewer immensely enjoyed reading the book. Since it is a set of notes it does not treat the subject exhaustively and gives us only glimpses of the author's attractive mathematical style. Yet it proceeds pleasantly and easily, and readily absorbs the reader's attention.

The reviewer would, however, like to take exception to the choice of subjects and certain aspects of their presentation. In spite of its name, this is not a book on Fourier transforms; it is a book on distribution functions studied with the aid of Fourier transforms. For example, the question of when a function is a characteristic function is barely discussed. To the best of the reviewer's knowledge the phrase "chance variable" or its equivalent is never mentioned. Yet the author is in effect compelled to introduce chance variables to clarify certain concepts. His section on "Examples" discusses mappings which are in effect chance variables of only moderate interest for the student of probability theory. There seems little point to so scrupulous an avoidance of the customary terminology of probability theory, with its attendant values of brevity and suggestiveness. All the "chance variables" which occur in this book are independent; this is often an unnecessary limitation. The question of the convergence of a series of independent chance variables is given considerable attention, yet neither the law of large numbers nor the central limit theorem are even mentioned.

The reviewer was disappointed by the extreme paucity of appropriate references. The zero or one law, which is due to Kolmogoroff (generalization by Paul Lévy) is wrongly ascribed by the author to Borel; the reference is *Rendiconti Palermo*, Vol. 29, 1909, where a special case of the strong law of large numbers is proved. The following minor remarks occur to the reviewer: (1) The proof (p. 168), that the distribution of Student's ratio is the same for all underlying joint distributions which are radially symmetric, can be considerably shortened. It suffices to note that Student's ratio is homogeneous of degree zero; the statement is really valid for all such functions, provided only that the functions are undefined on at most a set of probability

zero. (2) A simple proof by means of Fourier transforms of the fact that the totality of marginal distributions of every linear combination of k chance variables uniquely determines the joint distribution of the k variables (p. 160) is due to Cramér and Wold and to be found in the former's book, *Random Variables and Probability Distributions.*

The reviewer is of the opinion that, while no advanced worker in mathematics will want to be without a copy of this book, it is not particularly suited for statisticians or elementary students of probability theory. This is a great pity, and it is to be hoped that the distinguished author can be persuaded to treat the subject more comprehensively and systematically in a future book.

Random Normal Deviates: 25,000 Items Compiled From Tract No. XXIV (M. G. Kendall and B. Babington Smith's Tables of Random Sampling Numbers). *Herman Wold* (Professor of Statistics and Director of the Institute of Statistics, University of Uppsala, Uppsala, Sweden). University of London, University College, Department of Statistics, Tracts for Computers. London N.W.1: Cambridge University Press (Bentley House, 200 Euston Road), 1948. Pp. xiii, 51. Paper. 5*s*.

REVIEW BY H. BURKE HORTON
Bureau of Transport Economics and Statistics
Interstate Commerce Commission, Washington, D. C.

STATISTICIANS are indebted to Professor Wold for substantially increasing the number of random normal deviates available for general use. Prior to the publication of this set of 25,000 items there were in published form only 10,400 such items compiled by P. C. Mahalanobis from L. H. C. Tippett's random numbers.

In illustrating statistical processes a relatively small quantity of random normal deviates will usually suffice. However, for certain important research purposes, such as experimental deduction of the distribution function of a sampling statistic, large quantities of deviates are required. In this booklet Professor Wold has provided statisticians with 25,000 random normal deviates, each recorded to two decimal places. The items were derived from the Kendall-Smith tables of random numbers, column by column. It is unfortunate that Professor Wold's labor was doubled, due to the fact that use of the Kendall-Smith tables by rows yielded a set for which the variance was too large by a significant amount ($P = 0.7\%$). In the words of the author, "It is difficult to decide whether the failure with the first set [*by rows*] is accidental or due to some slight defect in the construction of Kendall-Smith's tables." If the latter possibility is the source of trouble, such difficulties may in the future be avoided, or at least minimized, by the use of recently developed convergent processes to generate the underlying set of random digits.

For the convenience of the user, sums and sums of squares for the fifty

items of each column are included. Tests for local randomness based upon sums, sums of squares, ranges, and sign runs, were applied to each page (500 items), to blocks of 5,000 items, and to the entire set. The results were in accord with the hypothesis of normality. Subsets yielding unusual test results are listed for the benefit of users of small sets from the table. For clarity, in line 2, page vii, substitute ".0100 and .9900" for "0 and 1".

Textual material accompanying the tables is clearly and concisely written. The author presents illustrated techniques for the construction of univariate, bivariate, and multivariate normal distributions, with specified underlying parameters. The material presented on the construction of a multivariate normal distribution is of particular value as a reference. This booklet will be a useful addition to the libraries of statisticians and statistical organizations requiring sizeable quantities of random normal deviates for research or educational purposes.

Say It With Figures. *Hans Zeisel* (McCann-Erickson, New York City). New York 16: Harper & Brothers (383 Madison Ave.), 1947. Pp. xix, 250. $3.00.

REVIEW BY GREGOR SEBBA
Professor of Economics, The University of Georgia
Athens, Georgia

THE CALAMITY that recently befell the public opinion polls points up the fact that the development of statistical methodology and techniques has by far outpaced the analysis of what Professor Paul F. Lazarsfeld, in his introduction to Hans Zeisel's book *Say It With Figures* aptly terms "the conceptional meaning of statistical procedures." Since Albert B. Blankenship reviewed the book in this JOURNAL (42: 666–7 D '47) from the point of view of commercial research only, it seems appropriate to discuss briefly its contribution to conceptual analysis and its usefulness to teachers of statistics.

Zeisel's book deals with three broad subjects: "Problems of Classification" (Part I), "Means of Numerical Presentation" (Part II) and "Tools of Causal Analysis" (Part III). Part I is primarily meant for users of the questionnaire and interview methods; Zeisel's discussion of "Don't know" answers (which can easily be adapted to "Undecided" answers) is particularly illuminating when applied to political opinion polls. But it is Parts II and III upon which the importance of the book rests. Among the "Means of Numerical Presentation," Dr. Zeisel singles out percentages and simple indices for a penetrating analysis of their logic. The use of per cent figures is not generally advisable but needs "specific justification" and can be decided upon "only with a complete background of concrete data and specific circumstances" (p. 72). Their use for comparing increase or decrease in two or more populations, in particular, is logically justified only if the change is (or is treated as being) "in exact proportion to the factors chosen as a base for per cent computation" (p. 80). It thus turns out that per cent comparisons

"offer only approximate solutions" since they merely discount "*a priori* the effects of concomitant variates" (R. A. Fisher); hence their use for comparison "will be justified [only] to the extent to which this a priori reasoning proves correct" (p. 81). In a two-dimensional table, per cents should be run in the direction of the variable to be studied for its effect, provided the sample is representative in this direction; if it is not, proper weighting becomes necessary (pp. 105–6). Of particular interest is Zeisel's subsequent study of the problem of reducing three- and more-dimensional tables (chap. 6), a problem arising because "only tables containing two variables can be presented in their entirety and still be clearly readable" (p. 127). There follows an illuminating discussion of simple indices of the type developed in sociometrics, leading up to the warning that "there is a certain danger that somewhere along the way from a clearly defined object to its mathematical symbols, the clarity of thought is lost; indices sometimes pretend to measure a concept which . . . turns out to be ambiguous and, therefore, not measurable. Neither a descriptive label nor an impressive mathematical formula are a safeguard against . . . indices which do not measure what they purport to measure" (p. 166).

Part III, "Tools of Causal Analysis," contains a superior treatment of cross-tabulation as a tool of research. Cross-tabulation "refines" and "explains"—though the explanation may turn out to be spurious; the distinction between "true" and "spurious" inter-variable correlation depending on whether or not the correlation reflects a direct causal connection, i.e., whether the explaining factor is asymmetrically or symmetrically connected with the two variables (p. 202). Answering the question when to cross-tabulate, the author sets down the rule that if a result is analyzed successively by various breakdowns and it is known or suspected that some of them are interrelated, then these interrelated breakdowns should be tabulated, not successively, but simultaneously (p. 203). A critical discussion of the panel technique of interviewing concludes the book.

Dr. Zeisel's study represents a step forward on the road indicated by such classical earlier treatises as Zizek's study of averages (1913), Winkler's monograph on relatives (1923) and Haberler's analysis of index numbers (1927). The book might be termed an essay in the logic of statistics procedures. The teacher of elementary and applied statistics will find in it useful numerical examples and charts of great forcefulness; and while much of the author's discussion is too refined for beginners, there remains enough to enable the teacher to go through the initial chapters of an elementary text without putting the students (and himself) to sleep.

Although Professor Lazarsfeld claims that "some of the inevitable errata" have been corrected in the 1948 printing, others seem to have passed unnoticed; among them an ugly blemish: in Table X-19 (pp. 240–2), there occur three questions and answers of the form: "If Hitler offered peace now . . . , would you favor or oppose such a peace?" Answer: "Yes."